

Plan for Analyzing and Testing Hypotheses

Preliminary Evaluation of the Learning Opportunities and Biological Consequences of Monitoring and Experimental Management Actions

PATH
**Preliminary Evaluation of the Learning Opportunities and Biological
Consequences of Monitoring and Experimental Management Actions**

Prepared by:

Calvin N. Peters and David Marmorek, (compilers / editors)
ESSA Technologies Ltd.
1765 West 8th Avenue, Suite 300
Vancouver, BC V6J 5C6

Contributors:

Jim Anderson
Nick Bouwes
Phaedra Budy
Tom Cooney
Rick Deriso
Al Giorgi
Rich Hinrichsen
Bill Muir
Charlie Paulsen
Charlie Petrosky
Steve Smith
Earl Weber
Rich Zabel

April 11, 2000

Citation: **Peters C.N. and D.R. Marmorek (compls./eds.)** 2000. PATH: Preliminary Evaluation of the Learning Opportunities and Biological Consequences of Monitoring and Experimental Management Actions. Prepared by ESSA Technologies Ltd., Vancouver, BC, 150 pp.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission from the PATH Planning Group (D. Marmorek, C. Peters, N. Bouwes, J. Geiselman, E. Weber, C. McConnaha, C. Toole).

Table of Contents

Table of Contents	i
List of Figures.....	iii
List of Tables	v
Executive Summary	ES-1
<i>ES.1. Background.....</i>	<i>ES-1</i>
<i>ES.2. Purpose of this Report</i>	<i>ES-2</i>
<i>ES.3. Description of the Experimental Management Model</i>	<i>ES-2</i>
<i>ES.4. Approach and Results</i>	<i>ES-6</i>
<i>ES.5. General Conclusions.....</i>	<i>ES-19</i>
<i>ES.6. Next Steps.....</i>	<i>ES-20</i>
1.0 Introduction.....	1
1.1 Background.....	1
1.2 Purpose of this Report	2
2.0 Experimental Actions.....	5
2.1 Base Case (continue 1978-1994 conditions).....	5
2.2 Modify Transportation, Measure changes in SARs.....	5
2.3 Turn Transportation On/Off, Measure D.....	6
2.4 Carcass Introductions / Stream Fertilization.....	7
2.5 Manipulate Hatchery Production	7
2.6 Natural River Drawdown of 4 Snake River Dams (A3)	10
3.0 Tools for Evaluating Actions.....	11
3.1 Introduction	11
3.2 EM Model Objectives.....	11
3.3 Model Outputs.....	11
3.4 Model Structure	13
3.5 Model Inputs	17
4.0 Results	29
4.1 Model Comparisons and Sensitivity Analyses.....	29
4.2 Evaluation of Generic Actions	37
4.3 Evaluation of Experimental Actions	45
4.4 General Discussion.....	58
5.0 Next Steps	63
6.0 References.....	65
Appendix A: Complete Descriptions of Actions	67
A.1 Continue Current Hydropower Operations and Estimate Post-Bonneville Dam Survival	67
A.2 Modify Transportation / Measure changes in SARs.....	72
Hypothesized Ratio = 1.2	75
Hypothesized Ratio = 1.0	75
A.3 Turn Transportation On/Off.....	76
A.4 Carcass Introductions / Stream Fertilization.....	89
A.5 Manipulate Hatchery Production	101
A.6 4-dam Drawdown.....	108
Appendix B. Bayesian Approach to Evaluating Learning.....	119
Appendix C. Reevaluation of the method used to predict SARs from Recruits/Spawner	121
Appendix D. A graphic contrast of hatchery steelhead abundance and spring chinook SARs for 1990 through 1995	123
Appendix E. Population Projections	131
Appendix F. Exploration of hypothesis tests of true and realized D values.....	133

Appendix G. Application of PATH retrospective analysis to assumptions in the stream fertilization experiment	141
Appendix H. Experimental Management of D	143
<i>Mathematics of SAR and D.....</i>	<i>143</i>
<i>Data 145</i>	
<i>Evaluating actions to optimize transportation.....</i>	<i>149</i>
Appendix I. Details of Bayesian and Bootstrap Sampling.....	152
Appendix J. Implementation of the Hydro Extra mortality hypothesis.....	153
Appendix K. Comparison of normal approximation to the actual distribution of estimated Δm's	154

List of Figures

Figure ES-1:	Estimation of m_t from historical data (left) and historical time series of m_t (right).....	ES-4
Figure ES-2:	Example future time series of m_t values for forward projections of $\Delta m=1,0$ on/off experiment. .	ES-5
Figure ES-3:	Method of estimating Δm from future time series of $\ln(R/S)$ data.....	ES-5
Figure ES-4:	Distribution of estimated Δm values for a 1/0 on/off type of experiment of various durations.	ES-6
Figure ES-5:	Probabilities of exceeding survival (left) and recovery (right) thresholds for various Δm values.	ES-7
Figure ES-6:	Probability of detecting three example effect sizes of Δm 's as the # of treatment years changes.	ES-8
Figure ES-7:	Probability of detecting three effect sizes after 20 years for generic on/off experiments	ES-10
Figure ES-8:	Confidence level that true D is > 0.705 at LGR (overall D is $> \sim .65$) for different future observed geometric means of D for a time series of given length.	ES-11
Figure 2-1:	Steelhead hatchery releases from Snake River hatcheries, 1957-1998.	8
Figure 2-2:	Schedule of experimental steelhead hatchery releases, assuming a 30-year experiment.	9
Figure 3-1:	SAR for delays in transport fish arrival Below Bonneville Dam.	19
Figure 3-2:	Change in total SAR by altering the fraction of fish transported.	19
Figure 3-3:	Patterns of SAR and smolts/spawner for Snake River wild spring/summer chinook.	21
Figure 3-4:	m_t vs. number of steelhead hatchery releases from Snake River hatcheries, 1957-1990.	24
Figure 3-5:	1995 Spring/summer chinook SAR vs. hatchery steelhead passage index: spring chinook PI.	25
Figure 4-1:	Time series of estimated m_t values, 1957-1994.	31
Figure 4-2:	Probabilities of exceeding survival (left) and recovery (right) thresholds for the 6 th best stock.	36
Figure 4-3:	Distribution of Δm 's as the # of treatment years changes.	39
Figure 4-4:	Probability of detecting three example effect sizes of Δm 's.	40
Figure 4-5:	Probabilities of detecting critical Δm values for various true Δm values & lengths of experiments.	41
Figure 4-6:	Confidence level that true D is > 0.65 at LGR for different future observed geometric means of D for a time series of given length.	48
Figure 4-7:	Year Effects (m_t 's) vs. PIT-tag SAR's. See text for details.	59
Figure A-1:	Survival Range/ Median Survival, 32 Tagging Sites, 1992-1998.	93
Figure A-2:	Regression of Snake River spring/summer chinook total extra mortality (including D).	102
Figure A-3:	Simulation of expected results for hatchery Experiment 1 from Peters (1999).	104
Figure A-4:	Simulation of expected results for hatchery Experiment 2 from Peters (1999).	105
Figure A.6-1:	Hypothesized change for Snake River stocks in differential mortality (μ), SAR, in-river survival (V_n) and upstream passage survival.	111
Figure A.6-2:	Hypothesized changes in μ and SAR for stream-type chinook from the Snake River and Upper Columbia in response to implementing A3 in 2003 and B1 in 2012.	113
Figure A.6-3:	Spawner to spawner ratios (S:S; natural log scale) for seven index stocks of Snake River spring/summer chinook, 1975-1993 brood years.	116
Figure D-1.	1990 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	125
Figure D-2.	1991 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	126
Figure D-3.	1992 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	127
Figure D-4.	1993 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	128
Figure D-5.	1994 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	129
Figure D-6.	1995 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.	130
Figure F-1.	Confidence level that true D is > 0.705 at LGR (overall D is $> \sim .65$) for different future observed geometric means of D for a time series of given length.	138
Figure F-2.	Needed geometric mean observed D to achieve different levels of confidence that LGR D is > 0.705 (overall D is $> \sim .65$), for data sets of different length.	139
Figure G-1.	Patterns of SAR and smolts/spawner (natural log scale) for Snake River wild spring/summer chinook, smolt years 1962-1994.	141
Figure G-2.	Smolt-to-adult return rates versus $\ln(\text{smolts/spawner})$, smolt years 1962-1992.	142
Figure H-1.	SAR for transport and inriver passing fish for 5 day intervals.	145
Figure H-2.	Yearling chinook Smolt passage index at LGR dam in 1995.	146
Figure H-3.	Yearling chinook D varies over season.	147
Figure H-4.	Illustration of transport SAR (dots), transport dam arrival distribution $f(x)$, Bonneville arrival distribution y_t , and distribution with a delay of δ	150

Figure H-5.	SAR for delays in transport fish arrival Below Bonneville Dam.	151
Figure H-6.	Change in total SAR by altering the fraction of fish transported.	151
Figure K-1.	Distribution of estimated Δm values (standardized to mean=0 and std. Dev.=1) for the 1/0 on/off generic experiment.....	154

List of Tables

Table ES-1:	Experimental management (ExpM) tasks of PATH.	ES-1
Table ES-2:	Δm equivalents of three example effect sizes decision-makers may be interested in estimating..	ES-6
Table ES-3:	Survival, recovery, and quasi-extinction metrics for the 1/0 on/off generic experiment.	ES-9
Table ES-4:	Summary of learning and biological results for 1/0 on/off generic action.	ES-9
Table ES-5:	# PIT-tags needed to detect if D (single year) < 0.65 given assumptions about the true value of D and the SAR for transported fish.	ES-11
Table ES-6:	Number of PIT-tags needed to detect survival ratio of 1.2 based on assumptions on the true survival ratio and the SAR of control fish.	ES-12
Table ES-7:	Summary of results for transport / no transport experiment.	ES-13
Table ES-8:	Summary of results for carcass introduction / stream fertilization experiment.	ES-14
Table ES-9:	Summary of results for hatchery reduction experiment.	ES-15
Table ES-10:	Δm values resulting from three sources of survival effects of drawdown.	ES-17
Table ES-11:	Summary of results for 4-dam drawdown.	ES-17
Table ES-12:	Summary of results for all actions.	ES-18
Table 1-1:	Experimental management (ExpM) tasks of PATH.	2
Table 3-1:	Alternative model assumptions.	15
Table 3-2:	Stock*spawner parameters for six Snake R. spring/summer chinook index stocks.	22
Table 3-3:	Carcass coefficients for Snake R. s/s chinook index stocks.	22
Table 3-4:	Δm series for hatchery action (upper bound).	24
Table 3-5:	Changes in system survival and Δm at equilibrium for each D assumption.	25
Table 3-6:	Example time series of changes in system survival and Δm	26
Table 3-7:	Δm for each D assumption.	27
Table 4-1:	Comparison of retrospective models.	29
Table 4-2:	Retrospective Results for Alpha-type Model (1957-1994).	30
Table 4-3:	Alpha-style (1957-1994) prospective results for the base case (1978-1994 conditions).	32
Table 4-4:	Probabilities of meeting escapement targets as Ricker-as increase.	34
Table 4-5:	Probabilities of meeting escapement targets as Ricker-as increase.	35
Table 4-6:	Average Ricker-as (1978-1994) and average initial spawners (95-99)	37
Table 4-7:	Series of Δm 's applied to spawner-recruit survival.	38
Table 4-8:	Results of generic action 1 ($\Delta m = 0/1$ in odd years).	39
Table 4-9:	Results of generic action 1 ($\Delta m = 1,0,1,0$ etc.).	41
Table 4-10:	Results of generic action 2 (allow Δm to vary around a mean value of 1).	42
Table 4-11:	Series of Δm for 0/1 5 yrs on/off	43
Table 4-12:	Results for generic action 4 (0/1 5 yrs on/off).	44
Table 4-13:	Results for generic action 5 (Delta-style model; 0/1 on/off).	44
Table 4-14:	Results for generic action #6 (0/1 for 10 years, then 1)	45
Table 4-15:	Number of PIT-tagged fish required in treatment and control groups in each year to ensure sufficient adult returns in each group.	46
Table 4-16:	Numbers of PIT-tagged fish required yearly in treatment and control groups to detect hypothesized levels of effects of the treatment under various assumed SARs (for control fish)	49
Table 4-17:	Results for transportation on/off action.	50
Table 4-18:	Control stocks the same for run of experiment.	51
Table 4-19:	Alternate treatment and control stocks.	51
Table 4-20:	Results of carcass introduction/stream fertilization action.	52
Table 4-21:	Results for hatchery action.	53
Table 4-22:	Results of drawdown actions.	54
Table 4-23:	Summary of results for all actions.	55
Table 4-24:	Time-series of year effects and SAR's from several PIT-tag based sources.	59
Table A-1:	Number of PIT-tagged fish required in treatment and control groups to ensure sufficient adult returns in each group.	70
Table A-2:	Numbers of PIT-tagged fish required yearly in treatment and control groups to detect hypothesized levels of effects of the treatment.	75

Table A-3: Operational and monitoring measures- transport-no transport adaptive management experiment	78
Table A-4: Variables to monitor/estimate for transport-no transport experimental management action.	79
Table A-5: Releases and Detections, Chinook Released and Migrating in 1998	81
Table A-6: “Seasonal” CJS Survival and Detection Estimates, Chinook Released and Migrating in 1998.....	82
Table A-7: Comparison of Goodness-of-Fit (R-Square) for Alpha-style Models With and Without Passage Model Offsets.	84
Table A-8: SAR’s and TCR’s for 1995 and 1996 Spring/summer chinook transport studies at Lower Granite (LGR)..	85
Table A-9: Example power analysis for a range of TCR and monitoring periods..	88
Table A-10: 32 Sites with tagging data and Number of fish tagged, 1992-98.	91
Table A-11: Site names, locations, and climate regions for 16 sites with 6 - 7 years of tagging data, 1992-1998.	92
Table A-12: Mean survival from tagging to LGR, 1992-1998.....	92
Table A-13: Annual average length of fish tagged and annual Palmer Drought Index (PDSI).	93
Table A-14: 5% and 95% naïve bootstrap confidence limits on survival from tagging to LGR, 1992-1998.....	94
Table A-15: Regression results, base case, weighted by $1/(\text{Survival Coefficient of Variation})$	96
Table A-16: Power of ability to detect additive survival increase, assuming no variation in treatment effect.	98
Table A-17: Power of ability to detect additive survival increase, with variation as noted in treatment effect.	99
Table A-18: Power of ability to detect additive survival increase, with variation as noted in treatment effect..	99
Table A-19: Observations and inferences for nutrient-driven stock viability hypothesis.	100
Table A-20: Hypothetical examples of two possible experiments to evaluate effects of hatchery steelhead production on Snake River spring/summer chinook salmon survival..	103
Table A-21: Example inference table for Hatchery-Caused Stock Viability Extra Mortality Hypothesis.	108
Table A-22: Example inference table for Hydro Extra Mortality Hypothesis.	117
Table F-1. D estimates from Bouwes et al. (1999) spreadsheet, wild spring/summer chinook	135
Table F-2. Weighted averages, geometric means, and standard deviations by project	135
Table F-3. Parameters for random draws of D by project, and expected value of D for each project from draw.	136
Table F-4. Geometric Mean LGR D vs. Geometric ‘Overall D ’ from 16 year time series of random draws of D	136
Table H-1. SAR data used to estimate parameters.	147
Table H-2. Calculated values of y_n , y_t , V_n , V_t , $D(x)$, $f(x)$ and $h(x)$	148
Table H-3. The adjusted differential delayed mortality DA according to eq(7) and R according to eq(8).....	149

Executive Summary

ES.1. Background

Experimental management is an explicit commitment to reducing key uncertainties that, because of their significance, are preventing the identification of better management policies. In experimental management, short-term experimental actions are used to learn about the system, and this information is used to guide decisions about long-term management actions. One of PATH's original objectives is to assess the ability to distinguish among competing hypotheses from future information, and advise institutions on monitoring, research, and experimental management actions that would maximize learning. Because we are concerned with ESA-listed salmon stocks, PATH recognizes that experimental management actions must both **maximize the ability to achieve conservation and recovery objectives** and **generate information to guide selection of better long-term management actions**. There is not universal agreement within PATH about the relative priority of these two potentially conflicting objectives.

In the PATH Final Report for Fiscal Year 1998, we set out a plan for evaluating experimental management actions (Table ES-1). The first three tasks in this plan are complete, and have resulted in the following short-list of actions for further evaluation:

- Modify transportation / measure D
- Transport / No Transport
- Carcass introductions / stream fertilization
- Manipulate hatchery production

In addition to these four experimental actions, we have also evaluated a base case, which assumes that 1978-1994 conditions would continue into the future, and natural river drawdown of four Snake River dams (A3). The base case is not an experiment. Some managers feel that 4-dam drawdown is a management experiment, while others are interested in what experimental actions can be done short of drawdown.

Table ES-1: Experimental management (ExpM) tasks of PATH.

Task	Task Description	Completed
ExpM1	Clarify ExpM approach recommended by SRP	✓
ExpM2	Describe ExpM options as variations to A1, A2, A3, etc.	✓
ExpM3	Detailed description of ExpM options with review from the PATH Scientific Review Panel (SRP), I.T., NWPPC	✓
ExpM4	Develop tools for quickly evaluating ExpM options	This report
ExpM5	Evaluate experimental management actions – effects on stocks versus amount of learning possible	This report
ExpM6	Evaluate experimental management actions across populations, including feasibility of implementation	
ExpM7	Using results from these evaluations, develop a research, monitoring, and evaluation plan to support the 1999 decision	

ES.2. Purpose of this Report

This report describes our progress toward implementing tasks ExpM4 and ExpM5 for the 6 short-listed actions. Our primary focus in the work accomplished to date has been to make a start at developing some tools and procedures for conducting quantitative analyses of experimental actions. We have developed a set of experimental management (EM) modeling tools that allow us to quickly assess the biological/conservation consequences and learning opportunities of actions that affect overall survival of Snake River spring and summer chinook salmon. These models are intended to provide a starting point for additional work after PATH is discontinued.

We have used these models to conduct some **preliminary** screening and analyses of the short-list of actions listed above. These analyses are preliminary because:

1. We have not done a thorough assessment of the feasibility of implementing these actions. Because of this, we have evaluated a set of generic and hypothetical experimental actions without speculating about how these actions might be actually implemented.
2. We have only looked at the effects of individual actions; combinations of actions may be more effective.
3. We assume that an action will have some effect, then assess the resulting biological and learning consequences. We have not assessed the weight of evidence in support or against the assumed magnitude of effects.
4. In most cases, we have only looked at how long it would take to detect effects in overall survival, from spawner-recruit data.

Our preliminary assessments should therefore be viewed as illustrations of “what if” scenarios of management experiments. We address the question “Suppose that a particular action could be feasibly implemented and had a particular effect on Snake R. spring/summer chinook populations: What would the biological consequences of such an action be, how difficult would it be to estimate that effect from spawner-recruit data with reasonable confidence, and what are the resulting trade-offs between learning and biological objectives?”. These assessments are useful for developing and testing our EM models, and for providing some broad guidance on the learning and conservation implications of various actions.

ES.3. Description of the Experimental Management Model

Outputs

A. *Biological*

The primary output of the model is projected numbers of spawners and recruits for seven Snake River index stocks of spring / summer chinook. From these, we calculate probabilities of exceeding 1995 BiOp recovery and survival thresholds¹ over 24 and 100 years (survival standards) and 24 and 48 years (recovery standards). We also calculate the probability of going to one spawner or less in a given year (over 10 and 100 years) as a quasi-extinction metric similar to that used by CRI in their August 1999 document.

¹ These are the probabilities that the number of spawners of 6 out of the 7 index stocks will exceed survival and recovery threshold numbers of spawners. Survival thresholds range from 150 to 300 spawners; recovery thresholds range from 350 to 1150 spawners, depending on the stock.

In order to calculate these metrics, we assume that actions will be maintained for the duration of each metric's time horizon (i.e. 24 and 100 years for survival probabilities, 24 and 48 years for recovery probabilities, and 10 and 100 years for quasi-extinction metrics). With the possible exception of the drawdown actions, this assumption is probably not realistic because if one discovers a suite of actions that meets survival and recovery requirements, one likely would not continue with the original on/off experiment. The population metrics included here may thus be viewed as a relative index of the biological consequences to the stocks, if the experimental actions were continued indefinitely.

Probabilities of exceeding survival and recovery thresholds are lower in this analysis than in previous PATH reports because of differences in some of the assumptions and data used in the model:

- Because we use 1978-1994 as representative of current conditions, we are assuming that the poor ocean conditions that existed in this time period continue into the future.
- We have assumed in most cases that extra mortality² is "here to stay". That is, we assume that the same high level of extra mortality that was experienced in 1978-1994 continues on into the future.
- This analysis uses updated spawner-recruit data which includes spawner data up to 1999. Spawner numbers in these years were generally low, with zero spawners in some years for Marsh Creek and Sulphur index stocks.

B. Learning

The main metrics of how much can be learned from an action are expressed in terms of the probability of estimating effects of an action over various time frames, or, conversely, how long it would take to estimate an effect with a certain level of confidence. Various criteria can be applied to determine how long an experiment needs to be run to estimate effect sizes that reflect the risk preferences of decision-makers. We present three examples for illustration:

- 1) one approach might be to require the experiment to not have a negative estimated effect on survival. In this case, decision makers would want to know the probability of estimating any non-zero effect on survival rates, and how this probability changes as the experiment goes on. This is the least stringent of the three examples; the effect can be estimated with high probability in a relatively short period of time.
- 2) decision-makers may want to know that the estimated effect of the action is close to (say, 80% of) its hypothesized effect. When hypothesized effects are large, this is generally the most difficult criterion to meet (i.e., probabilities of meeting it are lowest).
- 3) if one applies standard criteria for designing experiments, we would want to be fairly certain that we do not claim that an effect exists when in fact the action has no effect. To do this, we define a critical effect size (Δm^*), which is set at a level that minimizes the probability (0.05 or less) of incorrectly concluding that there is an effect when in fact there is none. The probability of detecting this critical effect size, if it exists, is called the "power" of the experiment; the higher this probability, the more "powerful" the experiment. This is the most difficult criterion to meet when hypothesized effects are small.

² Extra mortality is defined as any mortality occurring outside the juvenile migration corridor that is not accounted for by: (1) productivity parameters in the spawner-recruit relationship; (2) estimates of direct mortality within the migration corridor; (3) common year effects influencing both Snake River and Lower Columbia River stocks; and (4) random effects specific to each stock in each year.

Model Structure

The model is based on the Ricker model of Recruits vs. Spawners that is used for most analyses of Pacific salmon populations. Natural log units are used to linearize this model because this makes it easier to deal with the wide variability that characterizes most spawner-recruit data sets and to estimate the model's parameters³. The model can be expressed as:

$$\ln(R_{i,t}/S_{i,t}) = a_i + b_i S_{i,t} + m_t + \epsilon_{i,t}$$

or alternatively as:

R/S	\propto	productivity	carrying cap.	year	error term
		factor	factor	effect	
for		each stock,	each stock,	all stocks,	each stock,
		all years	all years	each year	each year

These parameters are estimated from historical spawner-recruit data, then used in forward projections to simulate the effects of actions. Assumptions about the effects of experimental actions are implemented in the model through the “ m_t ” or “year effect” term, which can be thought of a general survival factor for each year that affects all Snake River spring chinook stocks simultaneously. In the model, m_t values are calculated relative to the average survival rate from spawner to recruit over the entire historical time period (1958 to 1994). For years when $m_t = 0$, overall survival was equal to the long term average. When m_t is positive, overall survival was better than average; when m_t is negative survival was worse than average. Because m_t is in natural log units, every unit increase (decrease) in m_t increases (decreases) survival by a factor of 2.7 ($1 / 2.7$) relative to the historical average. For example, when $m_t = 1$, survival in that year was 2.7 times the historical average. When $m_t = 2$, survival in that year was 7.4X the historical average ($= 2.7 \times 2.7$). When $m_t = -1$, survival in that year was 0.37X the historical average ($= 1 / 2.7$).

Modeling Process

1. Estimate model parameters (a_i , b_i , m_t , $\epsilon_{i,t}$) from historical spawner-recruit data (Figure ES-1).

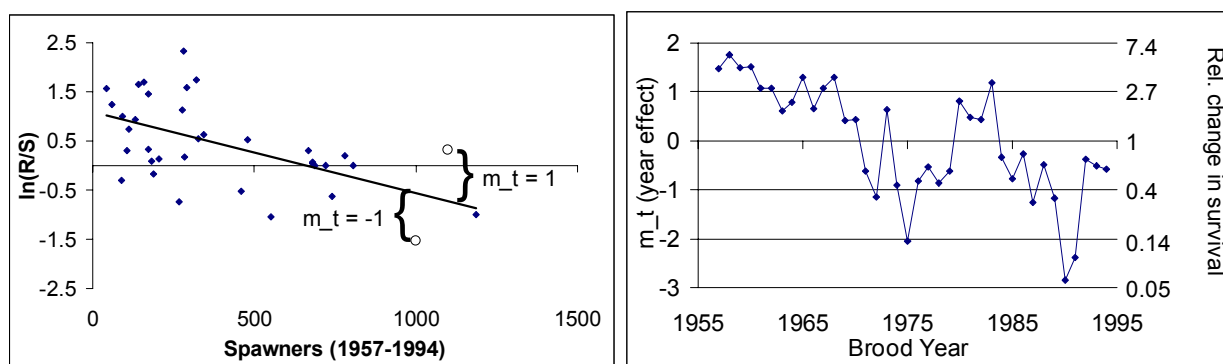


Figure ES-1: Estimation of m_t from historical data (left) and historical time series of m_t (right). Spawner-recruit data shown in left panel is a hypothetical dataset generated for illustration purposes.

³ When natural log units are used the error term, which for spawner-recruit data is assumed to follow a log-normal distribution, is transformed into a normally-distributed parameter. This allows us to fit a linear model to the log-transformed data.

2. Specify a future time series of m_t for simulating experimental actions. The future time series of m_t =

an historical m_t value selected at random from the 1978-1994 m_t values (this was used as the base period because conditions between 1978 to 1994 were assumed to be most like present conditions)

plus

a hypothesized effect on survival of the future action (this term is called Δm). For example, consider a hypothetical experiment in which some action is turned on and off in successive years. If this experimental action is hypothesized to cause a 2.7-fold improvement in survival in each year the action is implemented (“treatment year”) relative to years where the treatment is not applied (“control year”), the time series of Δm values for the forward simulation would be $\Delta m = 1, 0, 1, 0$, etc. for the duration of the experiment

The result is a future time series of Δm values that shows how an action is hypothesized to change overall spawner-recruit survival rates from the survival rates experienced between 1978-1994 (Figure ES-2 shows an example using the $\Delta m=1/0$ in on/off years example).

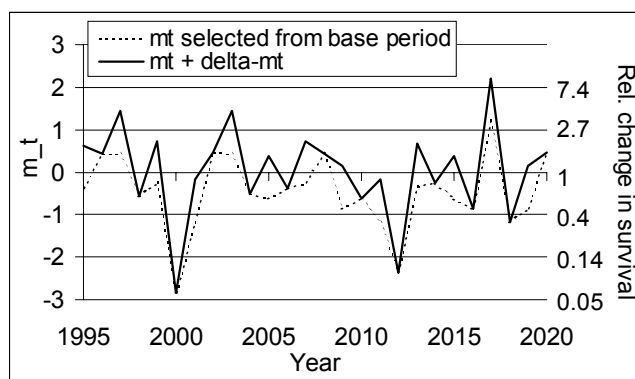


Figure ES-2: Example future time series of m_t values for forward projections of $\Delta m=1,0$ on/off experiment.

3. Use the future time series of m_t , along with historical estimates of the other model parameters (a , b , ϵ) to project populations through the experimental period. Simulate future data collection and analyses. Estimate Δm as the difference in average simulated $\ln(R/S)$ in treatment and control years (Figure ES-3). Calculate probabilities of recovery, survival, and extinction.

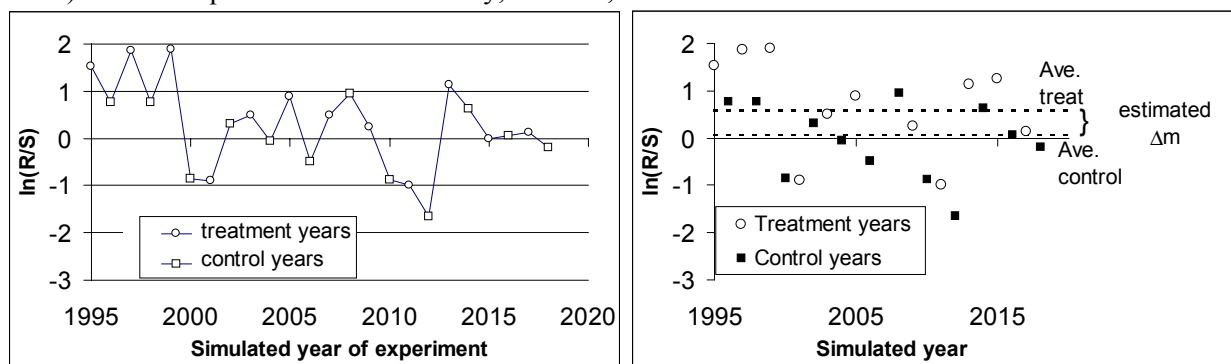


Figure ES-3: Method of estimating Δm from future time series of $\ln(R/S)$ data. Data are hypothetical examples for illustration purposes only.

4. Do this over multiple trials (i.e., many alternative futures) to get a frequency distribution of estimated Δm and biological metrics for different lengths of experiments (longer experiments = more data = better information) (Figure ES-4). Calculate probabilities of detecting various levels of Δm . Earlier in Section ES.3 we presented three examples of effects decision-makers may be interested in. These effect sizes can be translated into terms of Δm (Table ES-2). The frequency distributions are used to calculate the probabilities of estimating these Δm values.

Table ES-2: Δm equivalents of three example effect sizes decision-makers may be interested in estimating.

Effect	Corresponding Estimated Δm value
Experiment has no negative effect on survival	$\Delta m \geq 0$
Effect of the action is close to its hypothesized effect	$\Delta m \geq 0.8 \times \text{the "true" hypothesized } \Delta m \text{ value}$
Statistical "critical" effect size (Δm^*)	$\Delta m^* \geq 1.64 \times \text{std. deviation of the estimated } \Delta m$

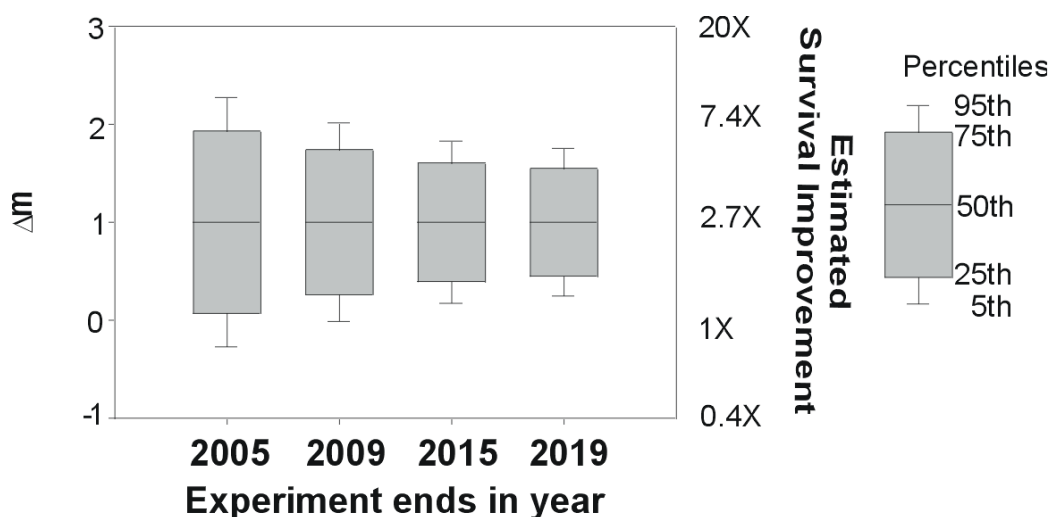


Figure ES-4: Distribution of estimated Δm values for a 1/0 on/off type of experiment of various durations.

Inputs

The primary model inputs are time series of Δm values for each experimental action, where these Δm 's represent hypotheses about how the action will affect overall survival rates relative to those experienced from 1978 to 1994. These are specified for a series of "generic" actions, in addition to the six experimental actions. Δm values and the model results for each action are described together in the next section.

ES.4. Approach and Results

Generic Actions

We looked at three sets of generic actions, in which various sizes and patterns of Δm values were projected into the future.

A. Constant Δm values

Various levels of Δm were held constant into the future (e.g. $\Delta m=0$ or 1 or 2, etc. in every year for 100 years into the future). The purpose was to conduct various sensitivity analyses of the model, e.g., to:

- Explore what level of annual survival improvement would be needed to meet 1995 BiOp survival and recovery standards;
- See how sensitive the survival and recovery probabilities are to the time periods over which they are measured (e.g. for 24-year survival probability: time period = 1996-2019 or 2000-2023); and
- Use as a surrogate for other actions by mapping their hypothesized Δm values to these generic results.

Results

Even a 7.4-fold increase in overall spawner-recruit survival (where this survival improvement is applied in each year of the simulation period) is not sufficient to meet the survival standard of 0.7 (Figure ES-5). A 2.7-fold improvement in survival is sufficient to meet the 48-year recovery standard of 0.5. Probabilities of exceeding survival and recovery thresholds are lower in this analysis than in previous PATH reports because of the differences discussed above.

Because of the recent low number of spawners, the 24-year survival probabilities are sensitive to which year the 24-year period starts. These probabilities are higher when 2000 is chosen as the starting year because the low spawner numbers between 1996 and 1999 are not included in the probability calculation. The 48-year recovery probabilities are not sensitive to recent spawner numbers because these probabilities are calculated over later simulation years (41 through 48). Results reported throughout the remainder of the report assume a starting year of 2000.

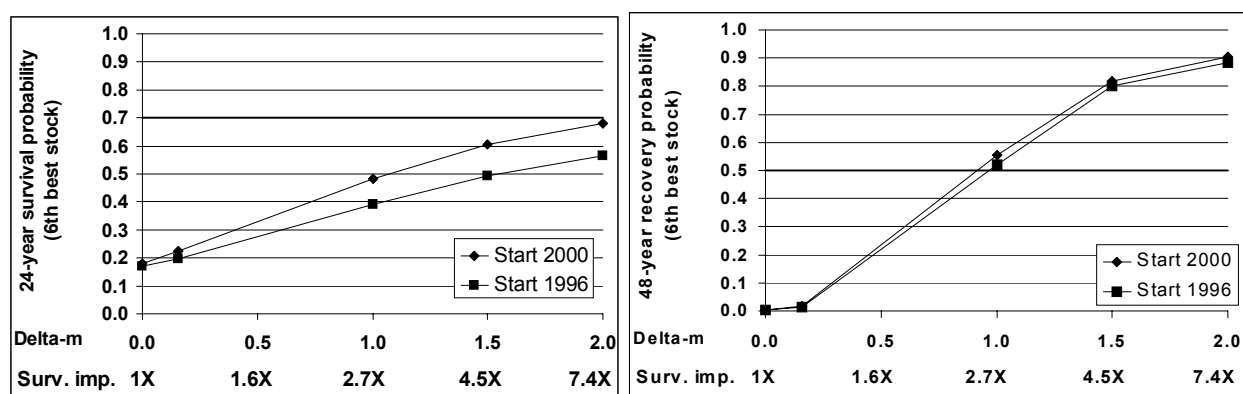


Figure ES-5: Probabilities of exceeding survival (left) and recovery (right) thresholds for various Δm values. Δm values are applied in every year of the simulation.

B. $\Delta m = 1, 0, 1$, etc. in on/off pattern

This set of runs implemented a $\Delta m = 1, 0, 1, 0$, etc. on/off pattern for varying durations starting in 2001 (Δm of 1=2.7-fold increase in survival). The purpose of running this set of generic actions was to explore model behavior, to provide a relatively simple example for explaining the approach and results, and to see in general how implementing treatments in an on/off pattern affects the ability to learn⁴. Altering

⁴ Different variations on this generic experiment were also explored, but produced similar results. See main report for details.

treatments in this way is expected to improve the ability to learn compared to holding Δm values constant (as in generic action set A) by reducing potential confounding with factors that happen to coincide with the start of the experiment in 2001.

Results – Learning Indicators

Precision of the Δm estimates improves as the duration of the experiment lengthens (see Figure ES-4). The gray box in that figure represents the range of Δm containing 90% of the estimated values. After only six years, there is a 90% chance that the estimated survival improvement will be between no improvement (relative to 1978-1994 average; i.e., $\Delta m = 0$) and a 7.4-fold increase ($\Delta m = +2$). However, after about 20 years, there will be a 90% chance that the estimated survival rate is between 1.6 ($\Delta m = 0.5$) and 4.5X ($\Delta m = 1.5$) the base case.

Probability of detecting the three effect sizes ($\Delta m = 0$, 0.8 of true, Δm^*) over time are shown in Figure ES-6. Decision-makers can use this graph to decide how long this experiment should run to achieve a desired level of certainty in detecting these effect levels. For example, the experiment should be run for 6 years if decision-makers want to be 95% confident that this action is at least doing no harm (i.e., has a 95% probability that Δm is at least 0). Or, applying the standard statistical criteria, one would need to run the experiment for 16 years to have at least an .8 probability of detecting the critical Δm value.

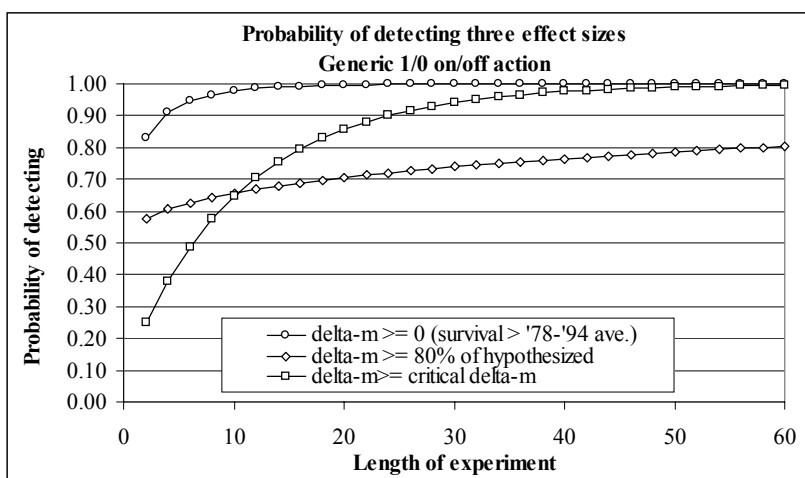


Figure ES-6: Probability of detecting three example effect sizes of Δm 's as the # of treatment years changes.

Results – Biological Indicators

We generally report results only for Sulphur Creek stock because in most cases this was the 6th best stock (Table ES-3).

Table ES-3: Survival, recovery, and quasi-extinction metrics for the 1/0 on/off generic experiment.

Metric	Value	Standard
24-Year Survival	0.35	0.7
100-Year Survival	0.51	0.7
24-Year Recovery	0.11	-
48-Year Recovery	0.15	0.5
10-Year Quasi-Extinction*	0.44	-
100-Year Quasi-Extinction*	0.65	-

* these metrics are insensitive to actions and are not presented further in the Executive Summary. See main report for details.

Results – Overall Summary

Table ES-4: Summary of learning and biological results for 1/0 on/off generic action.

"True" Δm ($\Delta_{\text{surv.}}$)	Year Exp. Ends	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of meeting	
					24-year Survival (Sulphur)	48-year Recovery (Sulphur)
1 (2.7X)	2009	0.98	0.66	0.65	0.35	0.15
	2013	0.99	0.68	0.76		
	2019	1.00	0.72	0.89		
	2029	1.00	0.76	0.97		

C. Various Δm values in in/off pattern

Approach

Various levels of Δm were implemented in an on/off pattern for 20 years, starting in 2001. For example, we implemented $\Delta m = 0.5, 0, 0.5, 0$, etc. or $\Delta m = 1, 0, 1, 0$, etc. for 20 years. The purpose of running this generic action was to explore how the size of the treatment effect influences the ability to estimate these effects.

Results

Figure ES-7 shows that larger effects are easier to estimate than smaller effect sizes. Note that when the true Δm is zero, there is still a 50% chance that one would estimate $\Delta m \geq 0$. That is, one would have a 50% chance of incorrectly concluding that the experiment had a positive effect on survival rate, when in fact it had no effect. One can reduce these types of errors by defining more statistically rigorous critical effect sizes (i.e. Δm^*). With this effect size, there is only a 5% probability of incorrectly concluding that the experiment had a positive effect on survival rate when the true effect was zero.

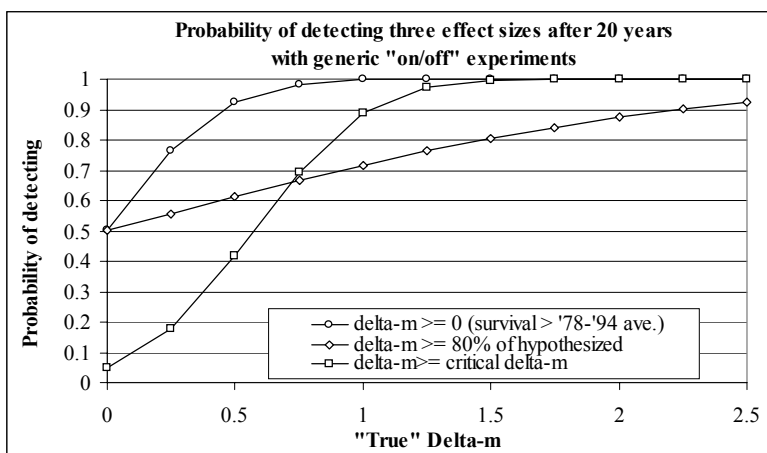


Figure ES-7: Probability of detecting three effect sizes after 20 years for generic on/off experiments with various Δm values.

1. Base Case (continue 1978-1994 conditions)

A. Approach

Maintain current operations, continue transportation studies, and monitor D (D is the ratio of post-Bonneville survival of transported fish to post-Bonneville survival of non-transported fish). Generate single year and multi-year estimates of D and see whether D is greater / less than hypothesized (e.g. $D < 0.35$ or $D > 0.65$).

This is a "base case" option with no major changes in operations or conditions. Overall survival rates are not expected to change from the 1978-1994 average (i.e., $\Delta m = 0$), because conditions and operations during this period of time are assumed to be representative of current conditions. Because there is no effect on spawner-recruit survival, the learning focus for this action is on sample sizes required to measure D within a given year and the number of years required to measure a multi-year average D with a certain level of confidence. The sample size calculations are based on the ability to determine whether D, measured in a single year, is greater than some hypothesized value. This calculation requires assumptions about the true value of D and the SAR of transported fish. The sample sizes provided (Table ES-5) represent the extreme high and low estimates of fish required for single-year experiments. Conducting this type of experiment over multiple years would increase the power to detect whether D is greater than a hypothesized value. Given an estimated multi-year average D (Dobs), the level of confidence that the true value of D is greater than a hypothesized value of 0.65 is presented as a function of time (Figure ES-8). It should be noted that the results presented in Table ES-5 and Figure ES-8 are based on separate analyses that make different assumptions on the variability of D.

B. Results - Learning Indicators

Table ES-5: # PIT-tags needed to detect if D (single year) < 0.65 given assumptions about the true value of D and the SAR for transported fish. This table provides the extreme high and low estimates of fish required for a single-year experiment.

Transport	Control*	Assumptions:
900,000	1.2 – 12 million	$D = 0.7$, SAR = 0.25%
3,350	6,700 – 67,000	$D = 1.0$, SAR = 2%

* sample size depends on reference group (larger if never-detected used as controls)

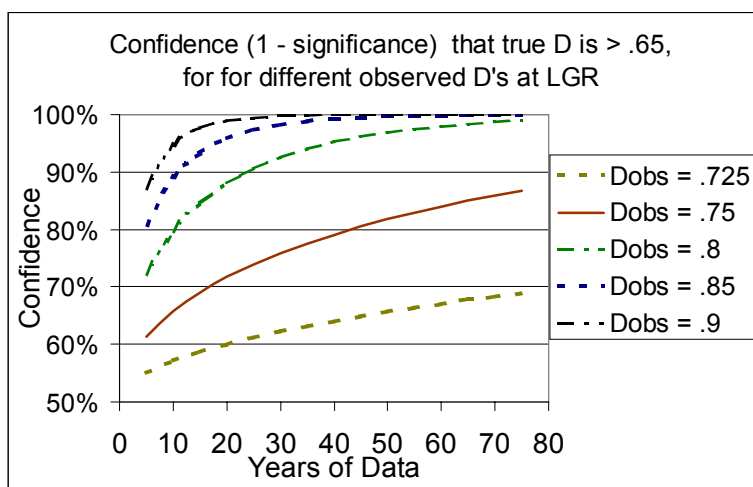


Figure ES-8: Confidence level that true D is > 0.705 at LGR (overall D is $> \sim .65$) for different future observed geometric means of D for a time series of given length.

C. Results - Biological Indicators

This action constitutes a base case scenario, where 1978-1994 operations and conditions are assumed to continue into the future. Under these assumptions, probabilities of exceeding survival and recovery thresholds are:

24-year survival (Sulphur) = 0.17	(Standard = 0.7)
48-year recovery (Sulphur) = 0.008	(Standard = 0.5)

D. Conclusions

1. Large numbers of PIT-tagged fish may be required to detect effects on SARs, depending on assumptions about future SARs and what groups are used as controls.
2. Assuming that 1978-1994 survival rates continue into the future, probabilities of exceeding recovery and survival thresholds are below the standards. Probabilities for all actions are lower than previous PATH analyses because assumptions about future climate and extra mortality are more pessimistic, and because recent spawner numbers for index stocks have been quite low.

2. Modify Smolt Transportation; Measure changes in SARs

A. Approach

We evaluated two suggested strategies for improving smolt transportation of spring/summer chinook:

- a) timing of delivery of smolts to estuary, and
- b) separate hatchery steelhead from wild chinook in barges.

Implementing these strategies could provide some information on the effects of improving smolt transportation in terms of SAR values, which in turn could be used to estimate incremental changes in D.

For both actions, we have assumed $\Delta m = 0.2$ (1.2-fold survival improvement) in every year. This is a relatively small Δm that would take a long time to detect in the spawner-recruit data. Therefore, our focus for measuring learning opportunities was on estimating the number of PIT-tagged fish needed to detect differences in survival of smolts that were transported separately from steelhead in barges from survival of smolts that were transported with steelhead (Table ES-6). The difference in survival is represented by a “survival ratio”, which is the ratio of the SAR of separated fish : SAR of non-separated fish. The sample sizes provided in Table ES-6 represent the extreme high and low numbers estimated.

B. Results - Learning Indicators

Table ES-6: Number of PIT-tags needed to detect survival ratio of 1.2 based on assumptions on the true survival ratio and the SAR of control fish. The sample sizes provided represent the extreme high and low numbers estimated.

Treatment (separated)	Control (not separated)	Assumptions:
3 million	2.4 million	True ratio = 1.25, SAR = 0.25%
1,200	2,400	True ratio = 2, SAR = 2%

C. Results - Biological Indicators

Assuming a $\Delta m = 0.2$ in every year, probabilities of exceeding survival and recovery thresholds are:

24-year survival (Sulphur) = 0.23	(Standard = 0.7)
48-year recovery (Sulphur) = 0.03	(Standard = 0.5)

D. Conclusions

1. Large numbers of PIT-tagged fish may be required to detect effects on SARs, depending on assumptions about future SARs and what groups are used as controls.
2. Modifying transport causes probabilities of exceeding recovery and survival thresholds to increase slightly from the base case, but not enough to meet the standards. This result assumes that the discussed modifications to transport result in a 1.2-fold increase in overall survival

3. Transport / No Transport

A. Approach

This is an on/off type of experiment where one would bypass and transport spring/summer chinook smolts in one year, then bypass and not transport in the next year. Alternating between these two strategies would continue for the duration of the experiment. The benefit of implementing such an approach would be to generate an estimable contrast in survival rates.

The survival difference one would expect depends on the transport:control ratio (T:C; the SAR of transported fish : SAR of non-transported fish). A T:C = 2 implies that SAR of transported fish is double that of non-transported fish. Therefore, one would expect to see overall survival rates in non-transport years that were about half the survival rate in transport years. This equates to a $\Delta m \approx -0.69$ in no transport years relative to transport years (transport years are used as the reference point because most fish during the base period of 1978-1994 were transported).

B. Results

Table ES-7: Summary of results for transport / no transport experiment.

"True" Δm ($\Delta_{\text{surv.}}$)	Year Exp. Ends	See Note 1			Prob. of meeting	
		Prob ($\Delta m \leq 0$)	Prob ($\Delta m \leq 0.55$)	Prob ($\Delta m \leq \Delta m^*$)	24-year Survival (Sulphur) Std. = 0.7	48-year Recovery (Sulphur) Std. = 0.5
-0.69 (0.5X)	2009	0.92	0.61	0.40	0.10	0.00
	2013	0.95	0.63	0.49		
	2019	0.98	0.65	0.63		

Note 1: These are prob. of being less than 0, ($0.8 \times \text{true}$), and Δm^* because effect is negative.

C. Conclusions

1. This experiment reduces the probabilities of exceeding the survival and recovery thresholds (relative to the base case) because SARs in non-transport years are reduced by half, assuming a T:C of 2. The survival and recovery results assume that the experiment is implemented in an on/off pattern for at least 24 (for survival) or 48 (for recovery) years.
2. If T:C is lower, the probability of exceeding the survival and recovery thresholds would be higher, but the probability of estimating this smaller effect would be lower.
3. After only 10 years, the experiment has a > 90% chance of estimating some effect, and > 60% chance of estimating 80% of the true effect. However, the power of the experiment does not meet the usual statistical criterion of 0.8, even after 20 years.

4. Carcass introductions / Stream fertilization

A. Approach

Introduce hatchery carcasses or add chemical fertilizers into rearing streams to increase nutrient levels and improve parr-smolt survival rates. Treatments would be varied by year and stream; the inclusion of treatment and control streams provides a spatial replicate that can be used to control for between-year variability. This action can be generalized to include any freshwater habitat improvement that can be applied to individual streams.

We specified a lower and upper bound for the hypothesized effect of this action on parr-smolt survival rate. The lower bound was that the action would have no effect (i.e., $\Delta m = 0$). This was based on a previous analysis of smolts/spawner data that suggested there has been no decrease in freshwater survival as the number of spawners has decreased since the 1960's. The upper bound assumes a 2-fold improvement in parr-smolt and spawner-recruit survival following carcass introduction/stream fertilization (equates to $\Delta m = 0.7$). This is based on an analysis of parr-smolt survival rates that suggests a positive correlation between the number of spawners and parr-smolt survival.

We looked at two alternative designs: one in which treatment and control stocks were the same every year, and one in which treatment and control stocks varied from year to year. We show results for Sulphur Creek and Poverty Flats stocks because in the experiments where treatment and control stocks were held constant, Sulphur was a treatment stock and Poverty was a control stock.

B. Results

Table ES-8: Summary of results for carcass introduction / stream fertilization experiment.

	"True" Δm	Year Exp. Ends	Prob (est. $\Delta m \geq 0$)	Prob (est. $\Delta m \geq 0.8$ of true)	Prob (est. $\Delta m \geq \Delta m^*$)	S=Sulphur (treatment) P=Poverty (control) Prob. of meeting	
						24-year Survival Std = 0.7	48-year Recovery Std. = 0.5
No effect (base case)	0.0		0.5	0.5	0.05	S = 0.17 P = 0.30	S = 0.008 P = 0.005
2X increase in parr-smolt; Same T,C	0.7	2010	1.00	0.80	0.99	S = 0.40	S = 0.30
		2020	1.00	0.86	0.99	P = 0.29	P = 0.00
2X increase in parr-smolt; Vary T,C	0.7	2010	1.00	0.83	0.99	S = 0.29	S = 0.08
		2020	1.00	0.91	0.99	P = 0.45	P = 0.10

C. Conclusions

1. The use of treatment and control stocks helps to control for factors that cause between-year variation in all stocks (e.g., climate conditions) and improves the probability of estimating effects. Assuming a 2-fold improvement in parr-smolt survival, these experiments are virtually certain to estimate some positive effect (> 0.9 probability of estimating $\Delta m > 0$) and have > 0.8 probability of estimating 80% of the true effect. The power of the experiment (probability of detecting a statistically significant effect) is far above the 0.8 criterion. Designs which vary treatment / control stocks have a greater probability of estimating 80% of the true Δm than designs that use the same treatment and control stocks.

2. Assuming that the action has the hypothesized upper bound effect, the probability of exceeding survival and recovery thresholds are increased substantially from the base case, but not enough to meet the standards. Varying treatment and control stocks may be a preferable design in terms of its biological consequences because probabilities of exceeding survival and recovery thresholds are increased for all stocks rather than just treatment stocks.

5. Manipulate hatchery production

A. Approach

The purpose of this experimental action is to explore the effects of hatchery interactions on overall survival of wild chinook by generating contrast in the number or timing of hatchery steelhead releases across years. In this example, we focus on reducing the number of smolts released by 25% and 50% in successive years. Current hatchery steelhead releases are around 12 million smolts; this hypothetical action reduces this number from 12 million in year 1, to 9 million in year 2, then to 6 million in year 3. This pattern of releases is continued in a 3-year pattern (12, 9, 6, 12, 9, 6, etc.) for the duration of the experiment.

We specified a lower and upper bound for the hypothesized effect of hatchery steelhead on survival rates of wild chinook. The lower bound hypothesis was that reducing hatchery releases would have no effect on wild chinook survival ($\Delta m = 0$). This hypothesis was based on an analysis that suggested no within-year relationships weekly SARs of wild chinook and the relative abundance of hatchery and wild fish at the dams. The upper bound assumes a negative, linear relationship between m_t estimated from the historical spawner-recruit data and the number of steelhead smolts released from hatcheries between 1958 and 1992. Based on this hypothesis, a 25% reduction in hatchery releases would produce a 2.1-fold improvement in survival ($\Delta m = 0.75$); a 50% reduction in hatchery releases would produce a 4.5-fold improvement in survival ($\Delta m = 1.5$).

B. Results

Table ES-9: Summary of results for hatchery reduction experiment.

"True" Δm (Δ surv.)	Year Exp. Ends	Prob (est. $\Delta m \geq 0$)	Prob (est. $\Delta m \geq 0.8$ of true)	Prob (est. $\Delta m \geq \Delta m^*$)	Prob. of meeting	
					24-year Survival Std.= 0.7	48-year Recovery Std. = 0.5
0.0 (no effect)		0.5	0.5	0.05	0.17	0.008
0.75 (2.1X) 25% reduction in hatchery SH	2005	0.84	0.58	0.26	0.41	0.26
	2011	0.91	0.61	0.39		
	2017	0.95	0.63	0.51		
	2020	0.96	0.64	0.56		
1.50 (4.5X) 50% reduction in hatchery SH	2005	0.98	0.66	0.27		
	2011	1.00	0.72	0.42		
	2017	1.00	0.76	0.54		
	2020	1.00	0.76	0.56		

C. Conclusions

1. The larger hypothesized effect size (i.e., 50% reduction in hatchery steelhead smolts) can be estimated in relatively short time period (6-12 years). It would take longer to estimate the smaller effect size that was hypothesized for the 25% reduction in hatchery releases. The probability of estimating a statistically significant Δm is below the 0.8 criterion for both effect sizes, even after 20 years.
2. The hatchery action, with 0, 25, and 50% reductions in hatchery steelhead releases in alternative years, can increase survival and recovery probabilities relative to the base case, but not enough to meet the standards. This result assumes that the hypothesized upper bound relationship between steelhead releases and spawner-recruit survival rate of wild chinook is true, and that the cycling between reductions continues for at least 24 years for the survival probability, and 48 years for the recovery probability.

6. 4-dam drawdown

A. Approach

The 4-dam drawdown action would breach four Snake River dams and stop transportation, while keeping hatchery production constant. Although there is disagreement over whether this is in fact an experiment, we evaluate it here for comparison. It is different from the other actions (except the base case and modify transport) in that there is no temporal or spatial contrast in effects: once the dams are breached, they are assumed to remain breached for the duration of the 100-year simulation period.

Effects on survival after drawdown come from three sources:

- Change in downstream passage survival and post-Bonneville survival of transported smolts (this depends on historical D assumptions). We used passage models to estimate these survival rates under base (1978-1992) and drawdown conditions.
- Change in extra mortality following drawdown. In addition to the “extra mortality is here to stay” or BKD hypothesis, we also examine the “hydro” hypothesis that says that extra mortality will revert to pre-dam (1957-1974) levels when dams are removed⁵. The change in extra mortality for this hypothesis was estimated by comparing estimated m_t values in the pre-1970 (pre-dam) period with estimated m_t values in the 1978-1994 period. This value also depends on the historical D assumption.
- Change in adult upstream survival. We assume a 15% improvement in upstream survival after drawdown.

Each of these hypothesized changes in survival can be expressed in terms of a Δm value (Table ES-10). The overall Δm is simply the sum of Δm values from each of the three sources.

⁵ This hypothesis is referred to as the Hydro II hypothesis in previous PATH analyses, and is described in the October 1999 PATH Experimental management Scoping Report and in Appendix H of the PATH Weight of Evidence Report.

Table ES-10: Δm values resulting from three sources of survival effects of drawdown.

D assumption	Δm due to Δ system survival	Δm due to upstream survival	Δm due to extra mortality		Combined Δm used in forward simulations	
			BKD	Hydro	BKD	Hydro
0.3	1.06	0.14	0	0.4 (1.5X)	1.2 (3.3X)	1.6 (5X)
0.6	0.53	0.14	0	0.93 (2.5X)	0.67 (1.9X)	1.6 (5X)
0.8	0.26	0.14	0	1.2 (3.3X)	0.40 (1.5X)	1.6 (5X)

B. Results

Table ES-11: Summary of results for 4-dam drawdown.

	"True" Δm ($\Delta_{\text{surv.}}$)	Year	Prob (est. $\Delta m \geq 0$)	Prob (est. $\Delta m \geq 0.8$ of true)	Prob (est. $\Delta m \geq \Delta m^*$)	Prob. of meeting	
						24-year Survival Std. = 0.7	48-year Recovery Std. = 0.5
D=0.3, 3-Year Delay BKD	1.2 (3.3X)	2010	0.99	0.69	0.79	0.41	0.56
		2015	1.00	0.75	0.95		
		2020	1.00	0.79	0.99		
D=0.3, 3-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.47	0.72
		2015	1.00	0.81	1.00		
		2020	1.00	0.86	1.00		
D=0.8, 8-Year Delay BKD	0.4 (1.5X)	2015	0.79	0.56	0.79	0.22	0.09
		2020	0.86	0.59	0.95		
		2025	0.90	0.60	0.99		
D=0.8, 8-Year Delay Hydro	1.6 (5X)	2015	1.00	0.74	0.95	0.35	0.72
		2020	1.00	0.81	1.00		
		2025	1.00	0.85	1.00		

C. Conclusions

1. If a low historical D is assumed, drawdown can meet the recovery standard but not the survival standard regardless of what is assumed about extra mortality. With a high historical D, drawdown meets the recovery standard with the hydro extra mortality hypothesis, but not with the BKD hypothesis.
2. There is a relatively good chance of estimating hypothesized effects because these effects are large and are applied in every simulation year (as opposed to the generic, transport on/off, and hatchery experiments, in which effects are applied in an on/off pattern). Probabilities of estimating both a positive effect ($\Delta m \geq 0$) and a statistically significant effect ($\Delta m \geq \Delta m^*$) are both high (≈ 0.8).
3. There is no cycling between treatment / control years in the drawdown action – once the dams are breached, they are assumed to remain breached for the duration of the simulation period. This increases the chances that measured effects may be confounded with other changes (such as changes in climate conditions) that are coincident with dam removal.

Overall Summary of Results

Table ES-12: Summary of results for all actions.

Action	"True" Δm	Year Exp. Ends	Prob (est. $\Delta m \geq 0$)	Prob (est. $\Delta m \geq 0.8$ of true)	Prob (est. $\Delta m \geq \Delta m^*$)	24-year Survival	48-year Recovery
Base case (1978-1994 conditions)	0.0		0.5	0.5	0.05	0.17	0.008
* Generic 0/1 on/off	1 (2.7X)	2009	0.98	0.66	0.65	0.35	0.15
		2013	0.99	0.68	0.76		
		2019	1.00	0.72	0.89		
		2029	1.00	0.76	0.97		
Modify Transport	0.2 (1.2X)					0.23	0.03
* Transport on/off (**)	-0.69 (0.5X)	2009	0.92	0.61	0.40	0.10	0.00
		2013	0.95	0.63	0.49		
		2019	0.98	0.65	0.63		
Carcass: No effect	0.0		0.5	0.5	0.05	Sulph. = 0.17 Pov. = 0.29	Sulph = 0.008 Pov. = 0.005
Carcass: 2X parr-smolt survival; treatment stocks constant	0.7 (2X)	2010	1.00	0.80	0.99	Sulph = 0.40 Pov. = 0.29	Sulph. = 0.30 Pov. = 0.005
		2020	1.00	0.86	0.99		
* Carcass: 2X parr-smolt survival; treatment stocks vary	0.7 (2X)	2010	1.00	0.83	0.99	Sulph = 0.29 Pov = 0.45	Sulph = 0.08 Pov = 0.10
		2020	1.00	0.91	0.99		
* Reduce hatchery production	0.0		0.5	0.5	0.05	0.17	0.008
	0.75 (2.1X)	2005	0.84	0.58	0.26	0.41	0.26
		2017	0.95	0.63	0.51		
		2020	0.96	0.64	0.56		
	1.50 (4.5X)	2005	0.98	0.66	0.27		
		2017	1.00	0.76	0.54		
		2020	1.00	0.76	0.56		
D=0.3, 3-Year Delay BKD	1.2 (3.3X)	2010	0.99	0.69	0.79	0.41	0.56
		2015	1.00	0.75	0.95		
		2020	1.00	0.79	0.99		
D=0.3, 3-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.47	0.72
		2015	1.00	0.81	1.0		
		2020	1.00	0.86	1.0		
D=0.8, 8-Year Delay BKD	0.4 (1.5X)	2010	0.79	0.56	0.79	0.22	0.09
		2015	0.86	0.59	0.95		
		2020	0.90	0.60	0.99		
D=0.8, 8-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.35	0.72
		2015	1.00	0.81	1.0		
		2020	1.00	0.85	1.0		

* these are on/off experiments

** probabilities are prob. of Δm being less than 0, 0.8 of true, critical Δm because these effects are negative

ES.5. General Conclusions

Biological

1. More than a 7.5-fold improvement in life-cycle survival is needed to meet the 24-year survival standard of 0.7.
2. A 2.7-fold increase in life-cycle survival is needed to meet the 48-year recovery standard of 0.5.
3. Survival and recovery probabilities in this analysis are lower than previous PATH results because:
 - assumes poor 1978-1994 ocean conditions continue
 - assumes extra mortality here to stay
 - uses updated spawner-recruit data
4. Using the hypothesized survival effects of the actions, all actions except transport on/off provide some survival improvement, but none meet survival standard (0.7). Only drawdown can meet the recovery standard (0.5), but this depends on D and extra mortality assumptions. Probabilities of exceeding survival and recovery thresholds for the transport on/off, carcass introduction (treatment and control stocks varied), and hatchery actions assume that these actions are implemented as on/off experiments for the duration of each metric's time horizon. This is probably not a realistic assumption because if an action appeared to be increasing survival it would likely be turned on permanently.

Learning

1. Most experiments have >0.8 probability of estimating some survival improvement (i.e., $\Delta m > 0$) within 5-10 years.
2. Actions that generate > 4 -fold survival improvement (i.e., some hypothesized responses to 4-dam drawdown and reductions in hatchery output) have about a 0.8 probability of estimating Δm of at least 80% of true value after 20 years. This is also true for actions that have smaller survival improvements but have spatial controls (i.e., carcass introductions/stream fertilization).
3. Actions that generate ≤ 2 -fold survival improvements with no spatial controls (i.e., transport / no transport, and some hypothesized responses to drawdown and hatchery reductions) have about a 0.6 probability of estimating Δm of at least 80% of true value after 20 years.
4. The probabilities of detecting a statistically significant Δm are low (i.e., less than the 0.8 criterion generally applied by statisticians) for all on/off experiments except for carcass introduction/stream fertilization. The use of spatial controls in that experiment improves the ability to estimate effects. These probabilities are high for drawdown because the hypothesized survival effects are large and are applied in every year, rather than in every other year as with the on/off type of experimental actions.
5. More complex designs / expanded monitoring of life-stage specific survival data is needed to improve the ability to detect effects. Spawner-recruit data is inherently "noisy" (i.e., between-year variation is large), and is affected by factors outside of direct management control such as climate and ocean conditions.

6. Wherever possible, within-year comparisons (e.g., treatment and control stocks for carcass introductions, treatment and control tag groups for hatchery/wild separation in barges) should be used to control for between-year variability and thus improve the ability of the action to estimate effects.
7. For status quo and modify transport options, large numbers of PIT-tagged fish may be required to detect effects on SARs, depending on assumptions about future SARs and what groups are used as controls. The largest estimates of tagged fish required may not be feasible.

ES.6. Next Steps

If further work on experimental management is undertaken after PATH ends, we recommend that effort be focussed on resolving / addressing the limitations of our preliminary analysis. Specifically, we suggest the following next steps:

1. Complete an assessment of the feasibility of implementing these experimental actions. For some actions, this will require consulting with regional management groups (e.g., hatchery managers, private and public landowners for carcass introductions/stream fertilization).
2. Assess the evidence in support / against our hypothesized effects of actions. The hypothesized values used in this report were suggested only as examples of values that might be used and approaches that could be used to develop hypotheses. Closer scrutiny of these and other hypotheses is needed. However, the hypothesized effects of most actions considered here are unlikely to be resolved without a series of well-planned experimental actions.
3. Use the model we have developed to explore alternative experimental designs and combinations of actions. There are many possible alternative designs to the ones we have used in our analyses, and many possible combinations of actions that could be explored (some of these combinations were discussed in the October 1999 Experimental Management Scoping Report). By strategically combining some of the experiments, one could test for multiple effects simultaneously.
4. Explore other monitoring to detect effects. Given the many factors that affect spawner-recruit data, and the large variability in spawner-recruit survival, the effects of actions on life-stage specific survival rates should be monitored in addition to the effects on spawner-recruit data. Such life-stage specific information may improve our ability to estimate the immediate effects of actions more precisely than the spawner-recruit data, although monitoring spawner-recruit data is still needed to assess overall survival responses.

1.0 Introduction

1.1 Background

Experimental management is an explicit commitment to reducing key uncertainties that, because of their significance, are preventing the identification of better management policies. In experimental management, short-term experimental actions are used to learn about the system, and this information is used to guide decisions about long-term management actions. These short-term experimental actions consist of deliberate changes to a system to provide contrast in treatments (Walters 1986), implemented in an experimental design that reduces confounding of management effects with other simultaneous events such as climate change. Large-scale management experiments often face challenges and limitations caused by a lack of suitable controls, lack of replicates, lack of baseline information, or difficulty in randomly assigning treatments to experimental units (some important traits of good experiments). In spite of these limitations, an experimental management approach produces a substantial improvement in the reliability and efficiency of information-gathering, compared to more passive management regimes (Walters 1986).

One of PATH's original objectives is to assess the ability to distinguish among competing hypotheses from future information, and advise institutions on monitoring, research, and experimental management actions that would maximize learning. Because we are concerned with ESA-listed salmon stocks, PATH recognizes that experimental management actions must both **maximize the ability to achieve conservation and recovery objectives** and **generate information to guide selection of better long-term management actions**. There is not universal agreement within PATH about the relative priority of these two potentially conflicting objectives.

In the PATH Final Report for Fiscal Year 1998, we set out a plan for addressing this objective (Table 1-1). Following consultation with the Implementation Team (I.T.) early in 1999, PATH established an Experimental Management Workgroup to more clearly define experimental management and generate a list of potential research, monitoring, and experimental management actions (i.e., the first three tasks in Table 1-1).

The identification and definition of potential monitoring and experimental actions was guided by the need to resolve two key uncertainties that had a large influence on PATH results: the magnitude of delayed effects of transporting smolts (the "D" value), and the incremental mortality experienced outside of the passage corridor by non-transported smolts ("extra" mortality). These uncertainties are not likely to be resolved directly by the experimental actions because D and (especially) extra mortality, and the factors that influence them, are difficult to measure empirically. However, they have provided a useful basis for identifying actions that are likely to have the largest effects on estimates of overall survival rates.

Table 1-1: Experimental management (ExpM) tasks of PATH.

Task	Task Description	Completed
ExpM1	Clarify ExpM approach recommended by SRP	✓
ExpM2	Describe ExpM options as variations to A1, A2, A3, etc.	✓
ExpM3	Detailed description of ExpM options with review from the PATH Scientific Review Panel (SRP), I.T., NWPPC	✓
ExpM4	Develop tools for quickly evaluating ExpM options	This report
ExpM5	Evaluate proposed experimental management actions – effects on stocks versus amount of learning possible	This report
ExpM6	Evaluate proposed experimental management actions across populations, including feasibility of implementation	
ExpM7	Using results from ExpM evaluation, develop a research, monitoring, and evaluation plan to support the 1999 decision	

The resulting list of ten candidate actions was described in the report “PATH: Scoping of Candidate Research, Monitoring and Experimental Management Actions (Working Draft)”, which was distributed in October 1999 to the I.T. and other regional policy groups. On Thursday November 4, PATH met with the Implementation Team to review the Experimental Management Report (Working Draft) and to get direction on priorities for future PATH activities. As a result of that meeting, PATH was directed by I.T. to implement tasks 4 and 5 in Table 1-1 (i.e., develop tools for evaluating actions; evaluate biological outcomes and learning opportunities) for the following short-list of actions:

- Modify transportation / measure D
- Transport / No Transport
- Carcass introductions / stream fertilization
- Manipulate hatchery production

In addition to these four experimental actions, we have also evaluated a base case, which assumes that 1978-1994 conditions would continue into the future, and natural river drawdown of four Snake River dams (A3). The base case is not an experiment. Some managers feel that 4-dam drawdown is a management experiment, while others are interested in what experimental actions can be done short of drawdown.

These six actions are described in more detail in Section 2 of this report.

1.2 Purpose of this Report

This report describes our progress toward implementing these tasks for the selected group of actions. Our primary focus in the work accomplished to date has been to make a start at developing some tools and procedures for conducting quantitative analyses of experimental actions. Given the emphasis placed on evaluating experimental management actions by the ISAB, the PATH SRP, and other analytical groups, we anticipate that the tools and preliminary analyses that we have completed to date can serve as a useful starting point for additional work after PATH is discontinued.

In this context, we have developed a set of experimental management (EM) modeling tools that allow us to quickly assess the biological consequences (measured in terms of NMFS survival and recovery standards, and CRI-type quasi-extinction metrics) and learning opportunities (measured in terms of the

precision of the estimate of the experimental effect) of any action that has an effect on overall survival of Snake River spring and summer chinook salmon. The EM models are flexible in that they can accommodate a range of input assumptions (e.g., which historical years are considered a relevant base period), can be run relatively quickly compared to other PATH models, and are relatively simple to make them more accessible to regional analysts. The models are described in Section 3 of the report. We have not had time to develop similar tools for Snake R. fall chinook, although the same general approach could also apply for that stock.

We have used these models to conduct some **preliminary** screening and analyses of the short-list of actions listed above. These analyses are preliminary because:

1. We have not done a thorough assessment of the feasibility of implementing these actions, which would include engaging regional groups in developing detailed plans for experimental actions. For example, we have not had the discussions with regional hatchery experts that would be essential for developing feasible hatchery actions. Because these assessments have not yet taken place, we have evaluated a set of generic and hypothetical experimental actions without speculating about how these actions might be actually implemented.
2. In this report, we assume that an action will have some effect, then assess how long it would take to detect that effect and how it would affect survival, recovery and quasi-extinction metrics. We have not assessed the weight of evidence in support or against the assumed magnitude of effects, although some of the actions were considered in the PATH Weight of Evidence Report.
3. We have only looked at the effects of individual actions; combinations of actions may be more effective.
4. In most cases, we have only looked at how long it would take to detect effects in overall survival, from spawner-recruit data. Although this information is useful for determining the effects of actions over the entire life cycle (i.e., both direct and delayed effects), these effects are more difficult to detect because of influences of ocean conditions and other factors that are outside of the management action. Ideally, one would also monitor survival rates over shorter life stages (e.g., SARs, parr-smolt survival) to detect more immediate effects of experimental management actions. However, we have only started to develop approaches for doing this.

Because of these limitations, our preliminary assessments should be viewed as illustrations of “what if” scenarios of management experiments. We address the question “Suppose that a particular action could be feasibly implemented and had a particular effect on Snake R. spring/summer chinook populations: What would the biological consequences of such an action be, how difficult would it be to estimate that effect from spawner-recruit data with reasonable confidence, and what are the resulting trade-offs between learning and biological objectives?”.

Although preliminary, our assessments are useful for a couple of reasons. First, evaluating the actions in this way provides a means for us to develop and test our EM models. These models can then be used to evaluate more detailed sets of actions after the limitations listed above have been addressed. Second, the evaluations provide some broad guidance on the learning and conservation implications of various actions. For instance, if an action is hypothesized to have a large effect on survival, and the analysis shows that that effect is not likely to be detected, then there may not be much point in developing detailed plans for actions that are likely to have even smaller effects. Results of these evaluations are described in section 4 of this report.

In an effort to keep the main section of this report as readable as possible, we have put most of the detailed technical analyses in Appendices while summarizing them in the main report.

2.0 Experimental Actions

This section summarizes some of the main features of the six actions evaluated in the report. Complete descriptions of the actions have been excerpted from the scoping document and are provided in Appendix A of this report. The “*Risks to Stocks*” discussed in each section refer to the additional risks of experimental actions relative to maintaining the status quo.

2.1 Base Case (continue 1978-1994 conditions)

Experimental Action: Continue transport evaluation studies in the Snake River using PIT tags for both yearling chinook salmon and steelhead. Conditions for in-river migrants would be optimized by maximizing spill at downstream projects during the migration.

Benefits and Amount of Learning Possible: Considering projections of potentially greater adult return rates in the next few years, another 5 years of marking large numbers of juvenile fish along with adult return data from fish PIT tagged in recent years and currently still in the ocean, answers to some large-scale questions are obtainable in 5 years. For example, if the mean annual value of D is actually 0.8, another 5 years of data will very likely allow us to rule out the value 0.35. To distinguish between mean values of 0.7 and 0.8, however, would take much longer. Appendix F presents some estimates of how long it would take to make such finer-scale distinctions with varying degrees of confidence. One would need to tag between 63,000 and 2 million smolts, depending on survival rates, the level of precision required, and what are used as control groups. These estimates are based on estimating D in a single year; other factors would have to be considered if the goal was to estimate a mean D across multiple years.

Risk to Stocks: If transportation and/or the hydropower system have large impacts on fish, continual operation of the hydropower system and transportation will increase the risks that stocks will not recover. Additional risks to stocks would be minimal since recent studies on spring/summer chinook have shown a benefit from transportation from Lower Granite Dam. Furthermore, by maximizing spill for in-river migrants, not all fish would be transported which would spread the risk between in-river migration and transportation as called for in the current Biological Opinion.

2.2 Modify Transportation, Measure changes in SARs

During the past couple of years, PATH participants have discussed various changes in methods of transportation that could potentially improve the survival of transported fish. In this report, we focus on two examples of actions that modify transportation of Snake River spring/summer chinook salmon:

A. Change arrival timing of transported smolts in the estuary

Experimental Action: We assume that changes in SAR are a result of estuary arrival timing. To improve the effectiveness of transportation, therefore, we can:

- a) alter the time required for transported fish to reach the estuary, or
- b) alter the daily fraction of transported fish

By these adjustments, the SAR experienced by arriving fish will depend on when they arrive in the estuary and by which passage route they take.

Benefits and Amount of Learning Possible: Clarifying the effects of ocean entry timing and interaction with other stocks during collection and transport may reduce the uncertainty about D and extra mortality for both transported and non transported fish. Between 22,000 and 6 million smolts will have to be

marked to estimate D , depending on the true value of D and the desired level of precision (Appendix A.1). See Appendix F for discussion of factors that affect estimation of D over multiple years.

Risks to Stocks: Efforts to improve survival of transported fish, using only experimental PIT-tagged fish, would most likely not increase additional risk to the general population. Survival improvements may not be sufficient to attain recovery or avoid extinction.

B. Separate hatchery steelhead from wild chinook in barges

Experimental Action: This action requires continuing the current PIT-tag transportation experiments but with some portion of the release groups composed of wild spring chinook tagged and transported to below Bonneville Dam in isolation of steelhead. SARs for both control fish (i.e., fish transported under current operations) and treatment fish (i.e., fish tagged and transported with reduced steelhead interactions) would be computed on a seasonal basis. Ideally, treatment and control groups could be transported in the same year, which would increase the power of the test by eliminating year-to-year variability. Treatment and control groups could be released throughout the season (on separate barges) on a randomized basis or on the same barge in separate compartments.

Currently, the Lower Granite Dam juvenile collection facility does not have the ability to separate fish by species or size. Building a new juvenile facility at Lower Granite Dam, to include separation capabilities, has been discussed for many years, but has not been completed due to lack of agreement on design and pending decisions on transportation and dam removal. The COE has funded the NMFS and the University of Idaho in recent years to evaluate potential separator designs including permanent primary separators and temporary secondary separators. Based on the results of these studies, either a permanent juvenile separator could be built at Lower Granite Dam or secondary separation methods employed within the existing bypass flumes, raceways, or both to separate wild yearling chinook salmon from larger hatchery steelhead. Temporary grading bars within the existing bypass flume leading to the transportation study raceways were successfully used to reduce handling of hatchery steelhead during marking for chinook salmon survival studies in past years.

Existing transport barges have separate compartments so that wild chinook salmon could be barged apart after separation and marking without greatly disrupting current transport operations.

Benefits and Amount of Learning Possible: Isolating the effects of interactions of hatchery stocks with wild fish during collection and transport will reduce the uncertainty about D and extra mortality for both transported and non transported fish. Also the action may result in new methods to increase the efficiency of transportation.

Risks to Stocks: The experimental actions outlined here would not increase additional risk to wild stocks since it is unlikely that co-mingling with hatchery steelhead provides any benefits.

2.3 Turn Transportation On/Off, Measure D

Experimental Action: Vary the intensity of transportation. In some years, most fish would be bypassed, dewatered, and transported, while in others nearly all fish would be bypassed but not dewatered or transported.

Benefits and Amount of Learning Possible: Because the experiment would alternate years when most fish are transported with years when (almost) none would be loaded into barges, it should be possible to

observe greater contrast in survival rates of transported and non-transported fish. This should greatly reduce the current uncertainties associated with the benefits (if any) of transportation.

The essence of this adaptive management experiment is to extend conventional, PIT-tag based experiments to include “true” controls, that would be more nearly representative of fish migrating in-river with little or no indirect influence of transportation. That is, in years with transport turned off, fish migrating inriver would be bypassed and primary-dewatered, but not secondary-dewatered. If secondary dewatering reduces subsequent survival, these “control” years should capture and measure the effect, whether it is evident from inriver survival downstream of transport projects, or in SAR’s.

Risks to Stocks: The obvious risk is that if transportation is beneficial, eliminating it for the run-at-large half of the time will increase mortality. On the other hand, if we had complete certainty about the effects of transportation, we would not carry out the experiment in the first place.

2.4 Carcass Introductions / Stream Fertilization

Experimental Action: Introduce salmon carcasses or introduce chemical fertilizers to increase stream nutrient levels. These actions can be thought of as representing any habitat improvement action that can be turned on and off, and can be applied in some tributaries but not others.

Benefits and Amount of Learning Possible: As nutrients increase, then parr-smolt mortality, and perhaps “extra” spawner-recruit mortality will decrease. Parr in about 30 rearing areas are already PIT-tagged, about 16 of which have data for six of the past seven years. The availability of spatial control rearing areas suggests that the power of this experiment to detect changes in parr-smolt survival could be quite high. For example, if 7 of the 16 sites are treated and 9 used as controls, power could range from 0.33 to 1.0 after only 3 years of the experiment, depending on the size of the actual effect on parr-smolt survival.

Risks to Stocks: Disease spread is possible if carcasses are used, and there may not be enough disease-free carcasses to conduct the experiment.

2.5 Manipulate Hatchery Production

Experimental Action: Manipulate Snake River hatchery steelhead production to reduce exposure of wild Snake River spring/summer chinook juveniles to levels at or below those experienced in the 1970’s. Exposure of spring/summer chinook juveniles to hatchery steelhead could be reduced by decreasing the number of steelhead smolts released, reducing the size of steelhead smolts at release, or delaying steelhead smolt releases until late in the migration season.

We use steelhead here as an index of hatchery releases for purposes of developing a set of numbers to this analysis. Possible mechanisms for the effects of hatchery fish on wild spring/summer chinook are documented in Submission 1 of the Weight of Evidence Report. One could also look at total hatchery releases (i.e., including hatchery spring chinook), but because the temporal pattern of total releases closely matches that of steelhead, the effects of proportional reductions in total releases are very similar to the effects of the same proportional reduction in steelhead releases.

Steelhead releases from Snake River hatcheries are currently around 12 million per year (Figure 2-1). Practical, legal, and other constraints on reducing this number experimentally are numerous. Consequently, developing actual experimental strategies for reducing steelhead releases (if such strategies are accorded a high priority by the region) will require extensive regional consultation. Because such

consultation has not yet occurred, the PATH experimental management workgroup has developed the following hypothetical hatchery experiment. We stress that this scenario is intended only as a demonstration of how one could evaluate experimental hatchery scenarios, once these strategies were developed through consultation. While our scenario conforms to some of the legal and practical constraints that have been voiced thus far, we recognize that there may be other difficulties that make this scenario infeasible.

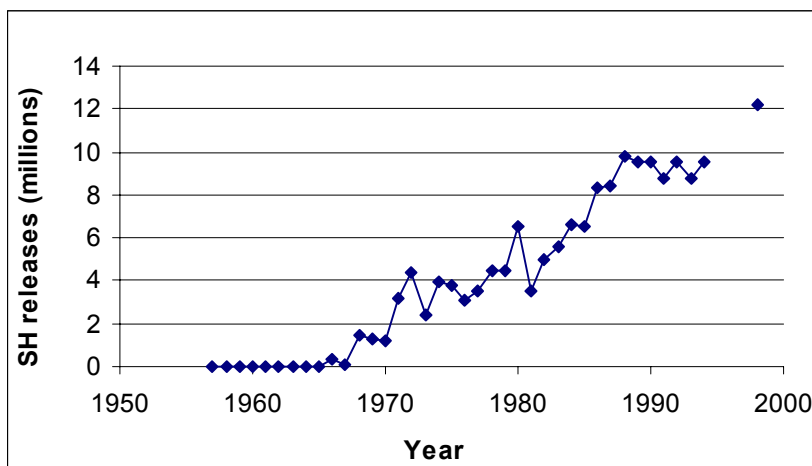


Figure 2-1: Steelhead hatchery releases from Snake River hatcheries, 1957-1998.

The hypothetical hatchery experiment evaluated here is based on three levels of annual releases:

high = twelve million releases, similar to the current 1998 level

intermediate = nine million, roughly what releases were during the late 1980's – mid 1990's

low = six million, a 50% reduction from current levels

Setting the high level at current releases recognizes the practical limits on increasing hatchery production capacity. Conversely, the lower level of 6 million represents a substantial reduction in production but does not encroach on existing conservation and restoration requirements. Determining the operational details of how such reductions could be effected will require consultation with regional hatchery experts.

These different levels of releases may alternatively be viewed as representing changes in the degree of overlap in migration of hatchery steelhead and wild spring/summer chinook smolts, either by delaying release of hatchery smolts or by separating hatchery and wild fish in barges (see Section 3.5.2). For instance, the low level of six million smolts could be used to represent the case where current levels of hatchery releases were delayed such that the overlap in migration with wild spring/summer chinook was reduced by half. In this case, although twelve million hatchery smolts are still released, only six million of them actually interact with wild spring/summer chinook migrants.

Temporal Pattern

The pattern of experimental hatchery steelhead releases should be planned such that maximum hatchery releases occur in years when smolts from the last maximum release year are produced. This ensures that adequate hatchery smolts are available for years when large releases are required. As an example, Figure 2-2 shows the three levels of releases implemented in a three year cycle. This three year cycle would be

appropriate, for example, for A-run type steelhead, which make up the majority of hatchery steelhead. These fish return as adults two years after leaving as smolts (one year spent in the ocean and one year overwintering in the river). Smolts are one year old when they migrate to the ocean. B-run type steelhead would have a four-year cycle (with two years spent in the ocean as adults before returning). These cycles could in theory be repeated indefinitely, and several experimental durations are evaluated in Section 4 of this report. In Figure 2-2, we have arbitrarily assumed a thirty-year experimental period (i.e., 10 experimental cycles).

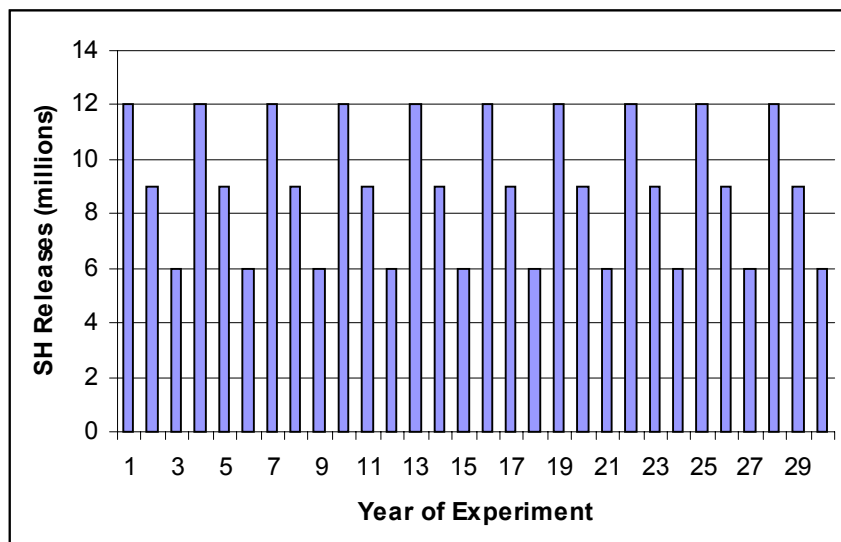


Figure 2-2: Schedule of experimental steelhead hatchery releases, assuming a 30-year experiment.

Benefits and Amount of Learning Possible: Determine (1) if there is support for the hypothesis that hatchery releases have affected extra mortality (and overall survival rate) of Snake River chinook, and (2) if reducing or eliminating exposure of wild Snake River spring/summer chinook migrants to hatchery steelhead can reduce total “extra mortality” of spring/summer chinook in the future, without breaching four Snake River dams. By simultaneously monitoring variables used to estimate D, and/or by simultaneously conducting transportation experiments, one could estimate the relative impacts of hatchery steelhead production on transported vs. non-transported spring/summer chinook (see Appendix A.1 and F for discussion of factors affecting estimation of D). The results of such a study could help determine which combinations of hydropower actions and hatchery management scenarios are most likely to result in achieving recovery goals for Snake River spring/summer chinook.

Risks to Stocks: Steelhead releases in the Snake River in 1998 totaled 12.2 million, of which approximately 3 million were used for conservation and/or restoration of native or local stocks. This leaves a possible maximum reduction in hatchery steelhead releases of 9.22 million without impacting conservation/restoration programs. Reductions should also consider the ability to maintain hatchery broodstock.

2.6 Natural River Drawdown of 4 Snake River Dams (A3)

Experimental Action: Breach Snake River dams, stop transportation, evaluate regional stock responses to help guide John Day drawdown decisions for listed Upper Columbia stocks. Hatchery production could be either pulsed or kept constant under this approach.

Benefits and Amount of Learning Possible: This is not an experimental action for Snake River drawdown decision; it is a long-term management action. However, implementation of this action would aid decisions on whether to restore natural river conditions in the John Day pool reach for listed salmon and steelhead in the Upper Columbia River. The staggered decision points for Snake River drawdown and John Day drawdown lend themselves to a staircase design, if implementation follows the same temporal pattern. Delaying Snake River actions while studies are conducted on John Day would negate this time step. Quantitative assessment of the probability of detecting effects should be determined in FY2000.

Risks to Stocks: According to the PATH FY98 and Fall Chinook Decision Analysis reports, 4-reservoir drawdown options (A3/B1) have the lowest risk, and highest biological benefits of any of the experimental actions proposed. Transportation-based actions had lower probabilities of meeting survival and recovery standards, and were less robust to uncertainties. The decision analysis indicates that recovery is generally likely for natural river options, regardless of which extra mortality hypothesis is correct. This approach would also help restore ecosystem function and benefit native lamprey, white sturgeon, and resident fish and wildlife, and non-listed anadromous stocks from above John Day pool.

3.0 Tools for Evaluating Actions

3.1 Introduction

We have developed two sets of models to evaluate experimental actions. The primary model is a simplified version of the existing PATH life-cycle model, and operates as a set of Fortran programs. These programs enable one to analyze the effects of changes in life-stage and/or life-cycle survival on probabilities of survival and recovery. All of the results in this report were generated by this primary set of models. The other model performs some simple power analyses, and operates as a set of Excel spreadsheets, along with an add-in regression package (XLSTAT). This set of spreadsheets is intended to allow analysts to better understand how the models work, and to give an easily accessible first-cut look at the power of various experiments to detect changes in life-cycle survival. Both sets of models are designed to accommodate a variety of life-cycle model structures and input data series.

3.2 EM Model Objectives

The primary objectives of the EM models are:

1. To express the amount of learning that is possible from each experimental action and combinations of actions, using metrics that are comparable across all actions.
2. To express the biological effects of taking the experimental actions, using metrics that are comparable:
 - across all experimental actions;
 - to other PATH analyses (i.e., survival and recovery standards); and
 - to other analyses (e.g., CRI probability of extinction).
3. The tool should be simple so that analyses can be completed in minutes rather than hours or days, and should be flexible and easy to use so that other analysts can run custom scenarios if they wish. The intention is to provide a tool that can be used in follow-up analyses of more detailed experimental actions even after PATH is discontinued.

3.3 Model Outputs

We have identified a set of primary and secondary outputs from the model (primary outputs are calculated directly in the model; secondary are calculated from primary).

3.3.1 Primary Outputs

Biological: Spawners, recruits, and other life-stage survival rates altered by the EM actions for seven Snake River index stocks of spring/summer chinook. These may include parr-smolt survival (for nutrient additions), in-river survival, and others. Originally, we had also intended to project SARs in the population model, based on an assumed relationship between SAR and R/S. However, we did not pursue this in this round of analyses because of certain problems with this approach (see Appendix C for details).

Learning: The main metrics of how much can be learned from an action are expressed in terms of the probability of estimating effects of an action over various time frames, or, conversely, how long it would take to estimate an effect with a certain level of confidence. Note that this is not exactly a “traditional” power analysis, because to estimate power one needs to specify a desired level of confidence and a

desired effect size that one wishes to detect. Various criteria can be applied to determine how long an experiment needs to be run to estimate effect sizes that reflect the risk preferences of decision-makers. We present three examples for illustration:

1. One approach might be to require the experiment to have a positive estimated effect on survival. In this case, decision makers would want to know the probability of estimating any non-zero effect on survival rates, and how this probability changes as the experiment goes on. In this case, decision makers would want to know the probability of detecting $\Delta m \geq 0$, and how this probability changes as the experiment goes on. This is the least stringent of the three examples; the effect can be estimated with high probability in a relatively short period of time.
2. Decision-makers may want to know that the estimated effect of the action is close to (say, 80% of) its hypothesized effect. This is a larger effect than just $\Delta m \geq 0$, so the probability of estimating it will be smaller. However, estimating this effect will give you greater confidence that the action is “working” (i.e., is having its hypothesized effect on survival). When hypothesized effects are large, this is generally the most difficult criterion to meet (i.e., probabilities of meeting it are lowest).
3. If one applies standard criteria for designing experiments, we would want to be fairly certain that:
 - a) we do not claim that an effect exists when in fact the action has no effect, and
 - b) if there is an effect, we will be able to detect it.

In most statistical applications, these general guidelines for designing experiments are quantified by requiring an experiment to have at least an 0.8 probability of detecting a “critical” value of Δm (Δm^*) that minimizes the probability (statisticians generally like this probability to be less than 0.05) of incorrectly concluding that there is an effect, when in fact the action has no effect. This critical value depends on the standard deviation of the distribution – the broader the distribution, the higher this critical value, and the lower the probability of detecting it. The critical value is calculated as $1.64 \times$ the standard deviation of the distribution of estimated Δm 's⁶. The probability of detecting this critical effect size, if it exists, is called the “power” of the experiment; the higher this probability, the more “powerful” the experiment. Using this approach, one can minimize the probabilities of making the two standard statistical errors: a) concluding that there is an effect when there is not one (Type I error); and b) failing to detect an effect that actually exists (Type II error).

3.3.2 Secondary Outputs

The probabilities of exceeding 1995 BiOp recovery and survival escapement thresholds⁷ are the primary conservation metrics produced by the model, to be consistent with previous PATH work. However, we also output the probability of going to one spawner or less in a given year as a quasi-extinction metric similar to that used by CRI in their August 1999 document. Time horizons for the survival and recovery standards are 24, 48, and 100 years; time horizons for the quasi-extinction metrics are 10 and 100 years.

To calculate these metrics, we must assume that experimental actions will be maintained for the duration of each metric's time horizon (i.e. 24 and 100 years for survival probabilities, 24 and 48 years for

⁶ The 1.64 value is based on a normal approximation of the true distribution of estimated Δm values. See Appendix K for a comparison of the normal approximation and the true distribution.

⁷ These are the probabilities that the number of spawners of 6 out of the 7 index stocks will exceed survival and recovery threshold numbers of spawners. Survival thresholds range from 150 to 300 spawners; recovery thresholds range from 350 to 1150 spawners, depending on the stock.

recovery probabilities, and 10 and 100 years for quasi-extinction metrics). With the possible exception of the drawdown action, this assumption is probably not realistic because if one discovers a suite of actions that satisfies survival and recovery requirements (however these are determined), one likely would not continue with the original on/off experiment. Instead, one would either decide on a “final” course of action or modify the action(s) and monitoring scheme(s) based on newly acquired information. The population metrics included here may thus be viewed as a relative index of the biological consequences of the experimental actions for the stocks, if these actions were continued indefinitely.

Probabilities of exceeding survival and recovery thresholds are lower in this analysis than in previous PATH reports because of differences in some of the assumptions and data used in the model:

- Because we are drawing base future m_t values from 1978-1994 m_t estimates, we are assuming that the poor ocean conditions that existed in this time period continue into the future. In previous PATH analyses, we assumed that the range of future climate conditions would be similar to that experienced between 1952 and 1990, which includes periods of both good and bad climate conditions.
- We have assumed in most cases that extra mortality⁸ is “here to stay”. That is, we assume that the same high level of extra mortality that was experienced in 1978-1994 continues on into the future. In previous PATH analyses, we had two alternative hypotheses: “hydro” (extra mortality goes away with improvements to the hydrosystem) and “regime shift” (extra mortality follows a 20-year cycle corresponding to climatic cycles).
- This analysis uses updated spawner-recruit data which includes spawner data up to 1999. Spawner numbers in these years were generally low, with zero spawners in some years for Marsh Creek and Sulphur index stocks. This essentially lowers the starting point for projecting future spawners, and makes it more difficult to exceed the survival and recovery spawner thresholds.

3.4 Model Structure

3.4.1 Overview

To evaluate experimental actions, we have developed a simplified form of the Ricker-type (density-dependent) population model used in previous PATH models. In general, we have tried to use the same assumptions in this model that were used in previous PATH modeling results, so that the two sets of results would be as comparable as possible. However, the model structure is designed to accommodate various alternative assumptions so that we can test the sensitivity of our model.

The population model is used in three steps to generate outputs:

1. The population model is fit to historical spawner-recruit data to generate estimates of model parameters (retrospective analyses).
2. The population model and historical parameter estimates, coupled with hypotheses about the anticipated effects of the experimental actions on overall survival rates, are used to project spawners and recruits. Estimates of the probabilities of quasi-extinction and of meeting survival and recovery standards are computed from the projections of spawners and recruits. The projection model is used to quantify potential learning from different experimental actions. We project many different possible spawners and recruits series for each action, then re-estimate the spawner-recruit model using the actual and projected SR data to obtain a sampling distribution of

⁸ Extra mortality is defined as any mortality occurring outside the juvenile migration corridor that is not accounted for by: (1) productivity parameters in the spawner-recruit relationship; (2) estimates of direct mortality within the migration corridor; (3) common year effects influencing both Snake River and Lower Columbia River stocks; and (4) random effects specific to each stock in each year.

the estimate of the experimental effect. The distribution of the estimated experimental effect is a measure of how precisely the experimental effect may be estimated after the experiment is completed.

3. Sensitivity analyses are used to understand the influence of various assumptions on the quasi-extinction, survival and recovery probabilities and the precision of the estimate of the experimental effect.

3.4.2 Population model

The model is an “Alpha-style” variant⁹ of the Ricker model:

$$\ln(R_{i,t}/S_{i,t}) = a_i + b_i S_{i,t} + m_t + \epsilon_{i,t} \quad [1]$$

where m_t = year-specific changes in Ricker-as over entire life-cycle, including passage mortality, extra mortality, year effects, harvest effects (depending on how one defines the $R_{i,t}$ terms; we define recruits ($R_{i,t}$) are defined as jacks+adults to the mouth of the Columbia R), etc.. These year-specific changes (year effects) sum to zero over 1957-1994. In the model, m_t values are calculated relative to the average survival rate from spawner to recruit over the entire historical time period (1958 to 1994). For years when $m_t = 0$, overall survival was equal to the long term average. When m_t is positive, overall survival was better than average; when m_t is negative survival was worse than average. Because m_t is in natural log units, every unit increase (decrease) in m_t increases (decreases) survival by a factor of 2.7 ($1 / 2.7$). For example, when $m_t = 1$, survival in that year was 2.7 times the historical average. When $m_t = 2$, survival in that year was 7.4X the historical average ($=2.7 \times 2.7$). When $m_t = -1$, survival in that year was 0.37X the historical average ($=1 / 2.7$). Natural log units are used because the error term for spawner-recruit data is assumed to follow a log-normal distribution. Therefore, using log-normal units transforms the error term into a normally-distributed parameter, and allows us to fit a linear model to the log-transformed data.

We used the same assumptions about in-river harvest and conversion rates for forward projections that were used in previous PATH models. We used current in-river harvest schedules for the Snake R. stocks, which are based on the escapement of the aggregate of all Snake River spring and summer chinook stocks. Conversion rates were selected randomly from the recent (1985 to 1999) values. Some modification to Equation [1] will be required for the carcass introduction / stream fertilization experiment. Because nutrient treatments would likely be applied to only a subset of the stocks, the model will have to distinguish between treated and untreated stocks. The error term ($\epsilon_{i,t}$) represents process and measurement error and follows a Gaussian distribution with mean zero.

Another variant of the life-cycle model, the “delta-style” model, was used in a sensitivity analysis. It partitions the m_t series above into common effects with lower river stocks (the delta series, δ_t) and the contrasts between the Snake and lower river stocks (the mu series, μ_t). The delta-style model is of the form

$$\ln(R_{i,t}/S_{i,t}) = a_i + b_i S_{i,t} - n_{i,t} X - \mu_t - \delta_t + \epsilon_{i,t} \quad [2]$$

⁹ This form of the model is most similar to the “alpha” model used in previous PATH analyses, in that there are assumed to be no survival effects that are common to both Snake River and downstream index stocks. Another variant of the model, the “Delta-style” model (described later in this section), assumes a common effect between upstream and downstream stocks and is thus most similar to the “Delta” model used in previous PATH analyses.

where $n_{i,t}X$ represents an effect of dams downstream of McNary Dam after 1970 (Deriso et al. 1996). SR data for brood years 1991-1994 were unavailable for the lower river stocks, so it was impossible to provide updates for the delta-style model. Our analysis suggests that there is an increase in precision of the experimental effect estimate when the delta-style model is employed instead of the alpha-style model described above (the SE decreases by about 20%). Using the model in this manner requires the assumption that the lower river stocks serve as controls for the Snake stocks.

3.4.3 Alternative model assumptions

There are several components of the model where different assumptions are possible, and one must make a choice between alternatives. Different choices may lead to quite different results, or may have no effect on results. The sensitivity analyses are designed to determine which of these choices affect the results, and which do not.

Components of the model where different assumptions may be appropriate, and the alternative assumptions that may apply, are summarized in Table 3-1. Model results using each of the alternative assumptions are compared in section 4 of this report.

Table 3-1: Alternative model assumptions.

"Data" - Related Assumptions	
Retrospective Period for model calibration	1957-94
	1952-1990 (Delta-style Model only)
"Method" - Related Assumptions	
Use Lower River stocks as "controls"?	Alpha-style - Don't include data from Lower River stocks
	Delta-style – Include data from Lower River stocks
Prospective Models Draw Year Effects from:	1952-1990 (Delta-style Model only)
	1978-1994
Error distributions for spawner and recruit projections include measurement error (Note 1)	Yes
	No
Parameter Distribution for forward projections (Note 2)	Bayesian Posterior
	Bootstrap

Note 1: The error term for the regressions, $\epsilon_{i,t}$, contains both process error and measurement error. For simulating future SR data (see Section 3.4.4), we had the option of shrinking the variance by 40%, to reflect the possibility of reducing this source of error in the future, or leaving the measurement error as a component of the error term. It turned out that the results of different experimental management designs were insensitive to changes in the assumed future measurement error.

Note 2: We explored two different methods of sampling from the parameter space for the forward model simulations: Bayesian sampling of the joint posterior distribution of the parameters (Gelman et al. 1995), and bootstrap sampling (Efron and Tibshirani 1993). Details of each approach are provided in Appendix I. It turned out that the bootstrap and Bayesian techniques yielded similar results, so Bayesian sampling was used throughout the report.

3.4.4 Forward projections

The EM models use the following process for conducting forward simulations. As noted above, the nutrient addition action, or any other action that affects only a subset of the stocks, would require some modification of the techniques (e.g., assume it increases the Ricker “a” for treated stocks but not for control stocks).

1. Estimate the parameters for Equation [1] a_i , b_i , and m_t from the historical spawner-recruit data (retrospective analyses)
2. Using Equation [1], the historical parameter estimates generated in Step 1, and assumptions about how an action will affect survival in the future, project populations through the experimental period. Assumptions about how an action will affect survival in the future are expressed in terms of a time series of $m^*(\text{exp})$ through the experimental period, where the $m^*(\text{exp})$ values are generated using:

$$m^*(\text{exp}) = m^*(\text{control}) + \Delta m \quad [3]$$

$m^*(\text{control})$ values used in forward projections will be selected from the series of historical m_t values estimated for years 1978-1990 (these years are assumed to be representative of current conditions). This is similar to the procedure used in previous PATH forward simulations where m and Δm were selected from the historical series and applied into the future. Δm values are input to the model, and represent hypotheses about changes to overall survival rates that are expected from experimental actions.

Two types of Δm values are investigated: a generic set (used to investigate general model behavior and responses), and an action-specific set (used to estimate the learning and biological consequences of the experimental actions). The input sets of Δm values are described in Section 3.5.

In some cases, it may be difficult or impossible to estimate Δm values for specific actions. For those actions, one can select a “proxy” set of Δm values from either a generic or another action’s specific set of Δm values. This proxy set of Δm values should approximate or bound the range of responses that might reasonably be seen from the action.

3. Each of the Δm scenarios will produce a time series of $m^*(\text{exp})$ (Equation [3]). Use the time series of $m^*(\text{exp})$ to project spawners and recruits over the experimental period, using model (1) above. Obtain an estimate of $\Delta m'$ (estimate of the experimental effect) from the simulated spawner-recruit series. The $\Delta m'$ values displayed in the results are the mean effects (i.e., treatment m_t 's – control m_t 's, over the duration of the experiment). A powerful experiment should produce a distribution of $\Delta m'$ that clusters tightly about the mean Δm used in the simulations, while a less powerful experimental will produce $\Delta m'$ that are more dispersed about the mean Δm . Calculate probabilities of recovery, survival, and extinction.
4. Do this over multiple trials, drawing from the frequency distributions of the estimates of a , b , retrospective m_t s, σ^2 , and the distribution of ϵ . The result of the multiple simulations will be a frequency distribution of the estimated experimental effect size, and distributions (means and 95% confidence intervals) of biological metrics. More sophisticated Bayesian approaches to evaluating how much is learned from an action are also possible (see Appendix B).
5. Sensitivity analyses revealed little difference between the jeopardy standards/quasi-extinction metrics or the precision of the experimental effect estimates when bootstrap sampling is performed instead of Bayesian sampling of the model parameters for model projections. Therefore we chose to use Bayesian sampling throughout.

6. The frequency distribution of the estimate of Δm was insensitive to removing the measurement error from the error term of the SR simulations.

3.5 Model Inputs

The primary input to the EM model is a time series of Δm values that represent hypotheses about changes to overall survival rates that are expected from experimental actions. Δm is calculated as the $\ln(\text{proportional change in survival})$. Thus if survival is hypothesized to double as a result of some action, Δm for that action = $\ln(2) = 0.69$. Two types of Δm values were specified: a generic set, and action-specific sets.

3.5.1 Generic survival improvements

The purposes of the generic sets of Δm values were to:

- a) investigate general model behavior, responses, and sensitivity to assumptions;
- b) provide a relatively simple example for explaining the approach and results; and
- c) see in general how implementing treatments in an on/off pattern affects the ability to learn. Altering treatments in this way is expected to improve the ability to learn relative to holding Δm values constant by reducing potential confounding with factors that happen to coincide with the start of the experiment in 2001. However, alternating between treatment and control years also means that the hypothesized survival improvements are only implemented in every other year, which will result in lower probabilities of exceeding survival and recovery thresholds.

We explored six generic actions:

1. $\Delta m=0$ until model year 2000, then $\Delta m=1,0$ in an on/off pattern (1,0,1,0, etc.) for 100 years.
2. *mean* Δm varies in an on/off pattern (0/1/0/1 ...) as in #1, but the treatment effect is drawn from a uniform distribution between 0.5 and 1.5 with mean one.
3. As in #1, but assumes measurement error eliminated.
4. $\Delta m=1,0$ in a 5-year on/5-year off pattern (1,1,1,1,0,0,0,0 etc.)
5. As in #1, but using a Delta-style life cycle model.
6. $\Delta m=1,0$ in an on/off pattern for 10 years, then 1 thereafter.

3.5.2 Base Case (continue 1978-1994 conditions)

Forward projections for this action represent the base case i.e., the results one would expect if 1978-1994 conditions were maintained indefinitely. In this case, the hypothesized effect of this action is zero, and Δm (which is designed to represent hypotheses about changes to overall survival rates that are expected from experimental actions) is also zero. Equation [3] thus reduces to:

$$m^*(\text{exp}) = m^*(\text{control}) \quad [4]$$

where $m^*(\text{control})$ values are selected each year from the 1978-1994 m_t values estimated from the historical spawner-recruit data. The base case, therefore, represents a future where the 1978-1994 years effects, drawn at random, are assumed to continue indefinitely. As such, it is a combination of previous PATH historical conditions (through brood year 1990), and more recent conditions that have not been explicitly modeled in previous analyses.

Note that in addition to describing “continue current operations”, this case also represents a case where an experimental action is taken but has no effect.

3.5.3 Modify transportation, measure changes in SARs

Change arrival timing of transported smolts in the estuary

There are two potential options for altering the arrival timing of smolts in the estuary:

- a) alter the time required for transported fish to reach the estuary, or
- b) alter the daily fraction of transported fish

To improve the effectiveness of transportation we can alter the time required for transported fish to reach the estuary, as defined by the factor d_t , or we can alter the daily fraction of transported fish as characterized by h . A relationship between these variables and the yearly averaged SAR is developed in Appendix H. Optimizing SAR then involves either altering the arrival time of transport fish into the estuary, which changes d_t or by increasing the percent of fish that are transported, which changes h . A third option of delaying the beginning of the transport season uniformly lowers SAR and so it is not considered further.

To explore effects of these two actions we can use the SAR distribution and the arrival time distribution of fish for 1995 (complete analysis is in Appendix H). The SAR is referenced to the time at arrival to Bonneville Dam. We assume the changes in SAR are a result of estuary arrival timing. We then adjust d_t and F (F is a fish condition factor that depends on when fish arrive at the transport dam) to alter the pattern over which fish enter the estuary. By these adjustments arriving fish experience SAR depend on when they arrive in the estuary and by which passage route they take. Our question then becomes “how would the average SAR for 1995 have been altered if we had moved fish at a different rate in transportation and if we had used a different transport schedule?”

The impacts of slowing barge transport by 5-day intervals on the overall SAR is illustrated in Figure 3-1 below. The impact of altering the percent of fish transported on the total SAR is illustrated in Figure 3-2. These relationships can be used as the basis for deriving a time series of Δm values for actions that affect these variables. As an example, Figure 3-1 shows that slowing barge transport by at least 5 days would have increased average SAR from 0.38 to 0.44%, an increase of about 20%. Assuming this change in SAR translates directly to a change in spawner-recruit survival, this equates to a Δm value of $\ln(1.2) = 0.2$. For illustration, this is the value we use in our evaluations in Section 4.

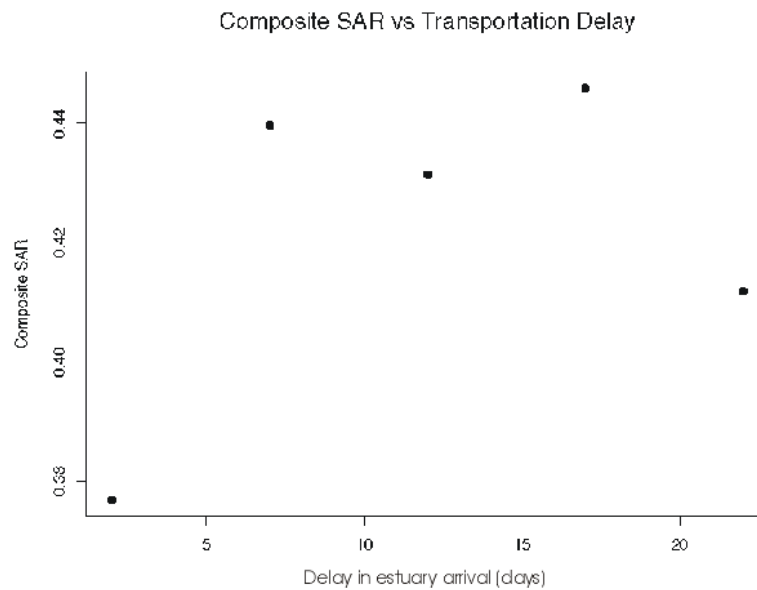


Figure 3-1: SAR for delays in transport fish arrival Below Bonneville Dam.

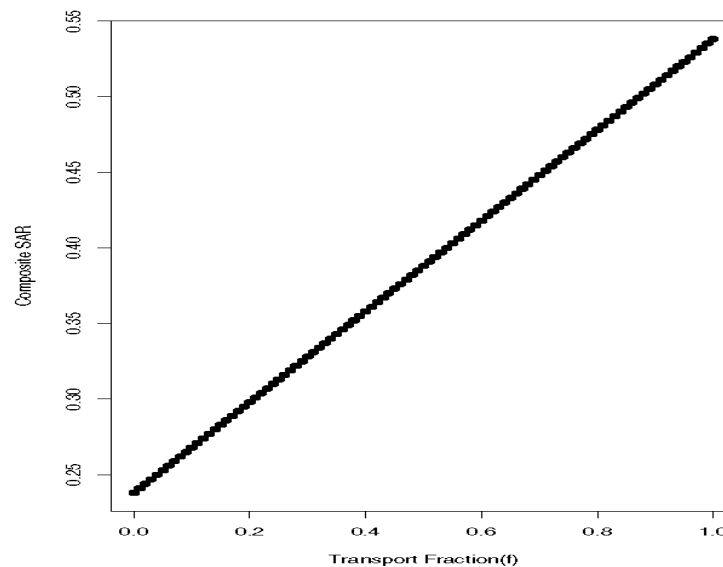


Figure 3-2: Change in total SAR by altering the fraction of fish transported.

Separate wild and hatchery fish in barges

There are several possible approaches for developing an estimate of Δm for this action. One approach would be to see if there is any relationship between the SAR of wild chinook and the ratio of hatchery steelhead releases: wild chinook releases in a given year. Another would be to use the regression between hatchery steelhead releases and m_t that is used to derive estimates of Δm for hatchery actions (see

Section 3.5.6), assuming that reducing the interaction between hatchery and wild smolts in the barges is equivalent to reducing the number of hatchery smolts released. For example, if the efficiency of separating fish in barges was 50%, one could use the estimated Δm for hatchery releases of 6 million smolts (50% of the current level).

Because we have not yet done these analyses, we make a simple assumption that separating hatchery steelhead from wild chinook smolts could increase SARs of wild chinook by a minimum 20% (based on the description of the action in Appendix A). This equates to a Δm of $\ln(1.2) = 0.2$. By coincidence, this is the same effect as the estuary arrival timing effect, and the two modify transport actions produce identical results. Therefore, we show only one set of results for the modify transport actions in this report.

3.5.4 Turn transportation on/off, measure D

Effects of transportation are modeled in terms of SARs to include both direct and delayed effects of transportation. Assuming a Transport: Control ratio (i.e., SAR of transported fish: SAR “control” or non-transported fish) of 2, this implies that SAR (and by extension, spawner-recruit survival) in years when fish are not transported will be half the SAR when fish are transported. This equates to a Δm value of $\ln(1/2) = -0.69$ in no transport years, relative to transport years (we use the transport years as the reference because under current operations most fish are transported). The negative Δm indicates that survival will decrease in years when fish are not transported, compared to years when fish are transported. Assuming an on/off temporal pattern, then, the time series of Δm would be 0 (transport year), -0.69 (no transport year), 0, -0.69 etc. for the duration of the experiment.

Assuming a lower T:C would mean less contrast in survival between transport and non-transport years and, consequently, would be harder to detect effects on overall survival. For example, a T:C of 1.2 would imply a Δm value of $\ln(1/1.2) = -0.18$.

3.5.5 Carcass introductions / stream fertilization

The October 1999 Experimental Management Scoping Report contains a number of references and a discussion of carcass/nutrient supplementation research on coastal coho, pink salmon, and steelhead. Unfortunately, no similar experiments have been attempted to date for any inland stocks or for any chinook stocks. This section contains brief summaries of three indirect lines of evidence regarding Snake River spring/summer chinook spawner abundance vs. estimates of smolts/spawner (3.5.5.1), carcass abundance and parr-smolt survival in the seven index areas (3.5.5.2), and Recruits/spawner vs. the abundance of carcasses (3.5.5.3). In sum, we use three hypothesized effects of carcass introductions based on these analyses: no effect ($\Delta m=0$); a small effect ($\Delta m=0.2$); and a large effect ($\Delta m=0.7$). The “always on” cases may be thought of as representing the “staircase” design developed in previous PATH experimental management reports.

3.5.5.1 Lower bound – stream fertilization/carcass introduction has no effect on survival

PATH retrospective analyses (Petrosky and Schaller 1998) do not provide any evidence of a temporal decrease in survival rate through the freshwater life stage that is proposed as the response variable in the experiments. This analysis (see Appendix G for complete analysis) indicates that while life-cycle survival rates and SARs decreased after completion of the hydrosystem, there was little evidence of decreased survival rates through the freshwater spawning/rearing life stage (Figure 3-3). Therefore, because the number of smolts produced per spawner did not decrease when the number of adult returns dramatically decreased, it seems unlikely that increases in carcass introductions will substantially improve spawner-to-smolt survivals. In addition, previous analyses of spawner recruit data through brood year 1990 detected no depensation (Deriso 1997), which might be expected if a carcass effect were present. In this case, Δm

resulting from a stream fertilization or carcass introduction action would = 0, which is equivalent to the “current operations” case described in Section 3.5.2.

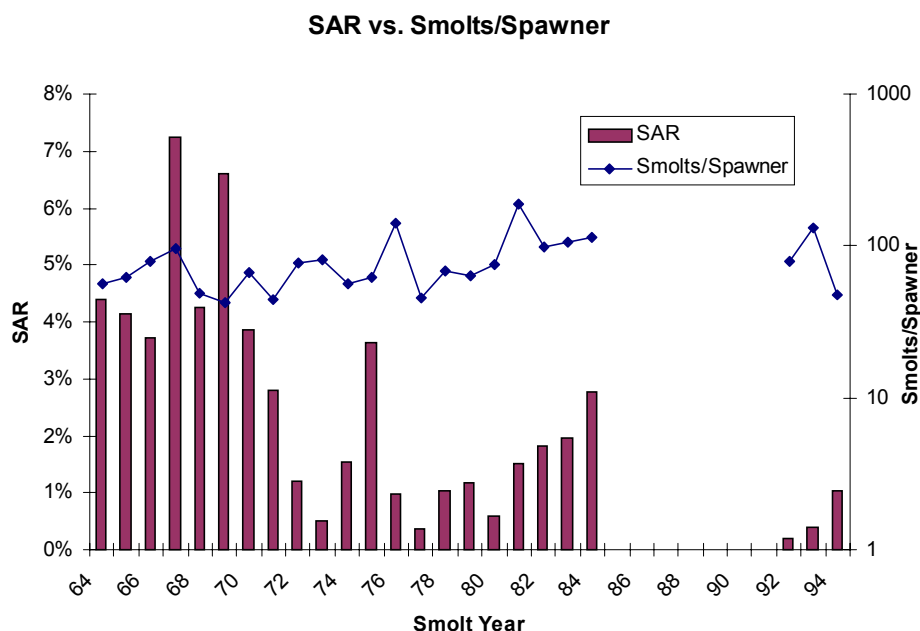


Table 3-2: Stock*spawner parameters for six Snake R. spring/summer chinook index stocks.

Stock	DF	Estimate	Std Err	Pr>Chi
BEAR/ELK	1	0.0003	0.0001	0.0027
IMNAHR	1	0	0	0.3681
MARSHC	1	0.0007	0.0002	0.0004
MINAMR	1	0.0001	0.0002	0.6501
Poverty Flat (SFS)	1	0.0001	0.0001	0.0564
SULFUC	1	0.0005	0.0007	0.4415

Obviously, the parameters are significantly different from zero for only three of the six stocks, and the significant parameters are small, averaging about 0.0003. On the other hand, if one were to add the nutrient equivalent of (say) 1000 spawners to each stream, this would in theory increase mean survival from about 0.30 (its mean for the stocks and years employed) to about 0.60 ($0.30 + 0.0003 \times 1000$). This is about a doubling of survival, which equates to a Δm value of $\ln(2) = 0.7$. One should regard these results with some skepticism, since there are few data points and little contrast in the spawner data. However, they do suggest that adding carcasses or nutrients may increase survival.

3.5.5.3 Evidence for carcass effects in spawner-recruit data

Again, in the absence of direct experimental evidence, a second indirect approach is to estimate a Ricker model that includes carcasses:

$$\ln(R_{j,t})/\ln(S_{j,t}) = a_j - b_j * S_{j,t} + m_t + c_j * S_{j,t+i} + e_{j,t} \quad [6]$$

Where j denotes stock, t denotes brood year, S denotes spawners, a_j is the Ricker “a” for stock j , “b” is the Ricker “b” for stock j , and “c” denotes a “carcass coefficient for stock j . By assumption, the “a” and “c” should be positive, while the “b” should be negative. Furthermore, as in the previous section, we assume that spawners for brood year $t+1$ provide food for the parr produced in brood year t , rearing in the subbasins in brood year $t+i$. The model is very similar to that used for most of the power analyses, with the addition of the c_j terms.

The coefficients of interest are the c_j 's. To estimate the model, we divided the 1957-1994 S/R data into two periods: 1957-1978 and 1978-1994. Only the first period produced results that were significantly different from zero. Whether this is simply chance or is due to much lower average spawner numbers in the 1978-1994 period is unknown. Results are shown in Table 3-3.

Table 3-3: Carcass coefficients for Snake R. s/s chinook index stocks.

	Estimate	Std Err	Pr>Chi
Bear	0.0001	0.0002	0.6733
Imnaha	0.0004	0.0002	0.0108
Johnson	0.0003	0.0004	0.4677
Marsh	0.0007	0.0003	0.0457
Minam	0.0006	0.0003	0.0362
Poverty	0	0.0002	0.9484
Sulphur	0.001	0.0005	0.0369

For the four of seven stocks with results significantly different from zero (Imnaha, Marsh, Minam, Sulphur), the average coefficient is about 0.0007, roughly $\frac{1}{2}$ the Ricker “b” values, and of the opposite sign. Note that for this data set, the S_t and S_{t+1} are positively correlated, Pearson “r” of about 0.51. If, as above, one were to add about 1000 carcasses per year, the results imply an increase in $\ln(\text{spawner} \rightarrow \text{recruit survival})$ (which is equivalent to Δm) of about 0.7 ($0.0007 * 1000$). This is consistent with the effect estimated from the analysis of parr-smolt survival. Again, the results should be viewed through skeptical spectacles, but in the absence of direct experiments they suggest that nutrient addition may be useful.

3.5.6 Manipulate hatchery production

Two analyses of hatchery release data place bounds on the effects of hatchery actions on survival of spring/summer chinook.

Upper Bound

An upper bound was based on a regression between historical estimates of m_t from spawner-recruit data and historical numbers of steelhead hatchery releases (Figure 3.-4). We use steelhead here as an index of hatchery releases for purposes of illustration for this analysis. Similar hypotheses could be developed about the effects of total hatchery releases (i.e., including hatchery spring chinook) on m_t , but this would not change the overall effect because the temporal pattern of total releases closely matches that of steelhead. Therefore, a proportional reduction in total hatchery releases would lead to the same Δm as the same proportional reduction in steelhead releases¹⁰. Obviously, though, the political, economic, and operational ramifications of a reduction in total hatchery releases would be much greater than reducing only steelhead releases.

Data on which the regression was based were from 1957 to 1990, the last year for which both hatchery releases and spawner-recruit data were available at time of writing. The regression was negative (lower survival at higher numbers of releases) and significant, explaining about 50% of the variability in the data. It is important to note that the fact that this regression exists does not constitute evidence that hatchery releases are the cause of reduced survival (i.e., correlation does not equal causation). In fact, such a correlation might be expected because hatchery releases were a mitigative measure implemented in response to declining fish populations. Therefore, the coincidence of increased hatchery production with declining survival rates does not necessarily mean that the one is the cause of the other. However, the regression in Figure 3-4 provides a convenient way to address the question “If hatcheries were the cause of declining survival, then what could we learn about this relationship by manipulating hatchery releases?”.

¹⁰ The regression equation for m_t vs. total hatchery releases is $m_t = -0.1144(\text{all_rel}) + 1.0223$ ($R^2=0.54$). Assuming total releases are now 24 million, a 25% reduction would give a Δm of 0.69; a 50% reduction would give a Δm of 1.4. These are similar to the values in Table 3-4.

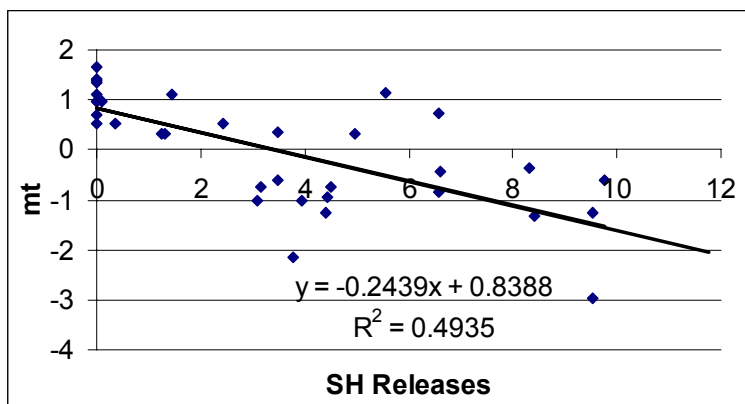


Figure 3-4: Regression of m_t vs. number of steelhead hatchery releases from Snake River hatcheries, 1957-1990.

This relationship is used to infer a Δm value during the forward projections of this experiment, depending on the level of hatchery releases during each year of the experiment. The Δm value is relative to current conditions, i.e., 12 million smolts released. Thus for years when 12 million smolts are released, Δm equals zero. When nine million smolts are released (a 25% reduction), $\Delta m \approx +0.75$, which is the difference between m_t at 12 million (-2) and m_t at 9 million (-1.25). $\Delta m \approx +1.5$ in years when 6 million smolts are released (50% reduction). For comparison, a Δm of 0.75 equates to approximately a two-fold increase in survival. $\Delta m = 1.5$ is approximately a 4.5-fold increase in survival. For the 3-year cycle of experimental hatchery releases described above, the time series of Δm used in forward projections is summarized in Table 3-4.

Table 3-4: Δm series for hatchery action (upper bound).

Year of Experiment	Δm
1	0
2	0.75
3	1.5
... repeat for duration of experiment ...	

Lower Bound

A lower bound on hatchery effects was based on a separate analysis that suggested no clear relationship between relative abundance of hatchery steelhead and spring/summer chinook (measured as passage indices) and SARs for spring/summer chinook from 1990 to 1995 (example for 1995 shown in Figure 3-5; full analysis in Appendix D). Based on these data, hatchery actions would have no effect on survival ($\Delta m=0$), which is equivalent to the base case described in Section 3.5.2.

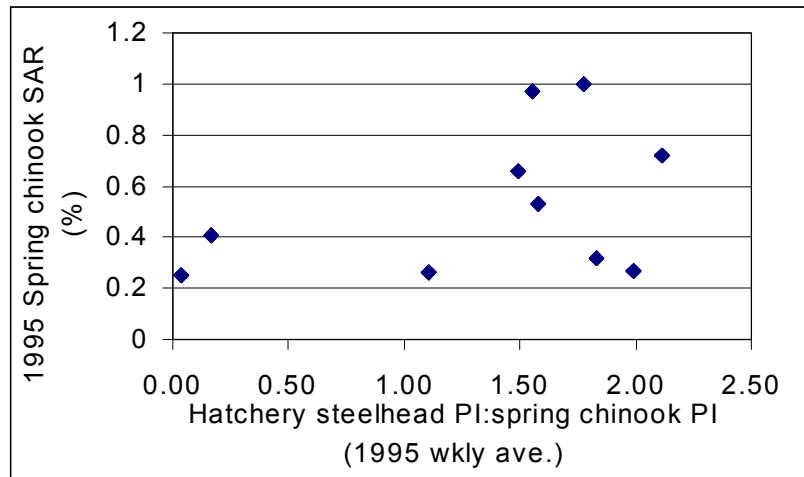


Figure 3-5. 1995 Spring/summer chinook SAR vs. hatchery steelhead passage index: spring chinook PI.

3.5.7 4-dam drawdown

Effects of 4-dam drawdown on m_t stem from three sources: changes in system survival, changes in extra mortality, and changes in upstream survival rate.

System survival

System survival is defined as

$$\text{Sys_surv} = e^{-M}(\text{DP}+1-P) \quad [7]$$

Where e^{-M} = total direct passage survival (a weighted average of transported and non-transported fish), P = the proportion of fish arriving below Bonneville that were transported, and D = the relative post-Bonneville survival of transported and non-transported fish.

Following drawdown, system survivals would change because fish are no longer transported (which affects P and e^{-M}), and because in-river survival of non-transported fish is increased (which also affects e^{-M}). In the EM models, changes are expressed as ratios of system_surv with drawdown: system_surv during the historical period 1978-1990 (i.e., when dams were in place). The change in system survival therefore depends on what is assumed about D (before drawdown) and in-river survival (both before and after drawdown). We estimated this change for three different D scenarios because the historical level of D cannot be resolved empirically: $D=0.3$, $D=0.6$, and $D=0.8$ (these correspond roughly to D hypotheses related to FLUSH passage model, CRiSP passage model, and NMFS PIT-tag analyses, respectively).

We used existing PATH passage model runs to estimate a change in system survival at equilibrium following drawdown. Changes in system survival at equilibrium for each D assumption is shown in Table 3-5.

Table 3-5: Changes in system survival and Δm at equilibrium for each D assumption.

D assumption	Change in system survival	$\Delta m = \ln(\text{change sys_surv})$
0.3	2.9	1.06
0.6	1.7	0.53
0.8	1.3	0.26

The time series of system survival changes will further depend on what is assumed about:

- the length of the pre-removal period (time between decision is made and construction begins). We used the same two assumptions as in previous PATH modeling: 3 years and 8 years;
- how quickly the unimpounded section of the river reaches equilibrium conditions. We used two assumptions: 2 years and 10 years, which is consistent with previous PATH results; and
- the length of the construction period (the amount of time it will take to remove the dams). We have assumed 2 years.

One example of a time series of changes in system survival rate is shown in Table 3-6. This example assumes an 8-year pre-removal period, followed by a 2-year construction period, followed by a 2-year transition period. Note that until construction is completed, the ratio of system_surv with drawdown: system_surv during the historical period =1 (i.e., there is no change in system survival until dams are removed). This example is for $D=0.3$.

Table 3-6: Example time series of changes in system survival and Δm .

Simulation Year	Change in system survival at equilibrium	$\Delta m = \ln(\text{change sys_surv})$ at equilibrium
1 (pre-removal)	1.0	0.0
2 (pre-removal)	1.0	0.0
3 (pre-removal)	1.0	0.0
4 (pre-removal)	1.0	0.0
5 (pre-removal)	1.0	0.0
6 (pre-removal)	1.0	0.0
7 (pre-removal)	1.0	0.0
8 (pre-removal)	1.0	0.0
9 (construction)	1.0	0.0
10 (construction)	1.0	0.0
11 (transition)	2.0	0.69
12 (equilibrium)	2.9	1.06
13 (equilibrium)	2.9	1.06
14, etc.	2.9	1.06

Extra Mortality

Extra mortality is defined as any mortality occurring outside the juvenile migration corridor that is not accounted for by: (1) productivity parameters in the spawner-recruit relationship; (2) estimates of direct mortality within the migration corridor; (3) common year effects influencing both Snake River and Lower Columbia River stocks; and (4) random effects specific to each stock in each year. There are three hypotheses about effects of drawdown on extra mortality: “BKD”, “regime shift”, and “hydro”. We focus on the BKD and the hydro hypotheses because these provide a lower and upper bound (respectively).

BKD

The BKD or “here to stay” hypothesis says that extra mortality will be unaffected by drawdown. Therefore, the Δm resulting from effects of drawdown on extra mortality will be 0.

Hydro

The hydro hypothesis says that extra mortality will revert to pre-dam (1957-1974) levels once dams are removed. This hypothesis is referred to as the Hydro II hypothesis in previous PATH analyses, and is described in the October 1999 PATH Experimental management Scoping Report and in Appendix H of the PATH Weight of Evidence Report. A description of how the hydro hypothesis was implemented in this model is provided in Appendix J.

Upstream survival rate

The effect of drawdown on upstream survival rates was estimated by comparing the average pre-dam (pre-1970) upstream survival rates to average post-dam (1976-1990) survival rates. The average increase was 15%, which equates to a Δm value of $\ln(1.15) = 0.14$. This increase is assumed to take effect immediately after construction is completed; there is no transition period.

Combined effect of change in system survival, extra mortality, and upstream survival rate

The combined effect of changes in these two components is simply the sum of their Δm values. Combined Δm values at equilibrium for each of the D assumptions are summarized in Table 3-7.

Table 3-7: Δm for each D assumption.

D assumption	Δm due to Δ system survival	Δm due to upstream survival	Δm due to extra mortality		Combined Δm used in forward simulations	
			BKD	Hydro	BKD	Hydro
0.3	1.06	0.14	0	0.4 (1.5X)	1.2 (3.3X)	1.6 (5X)
0.6	0.53	0.14	0	0.93 (2.5X)	0.67 (1.9X)	1.6 (5X)
0.8	0.26	0.14	0	1.2 (3.3X)	0.40 (1.5X)	1.6 (5X)

4.0 Results

4.1 Model Comparisons and Sensitivity Analyses

4.1.1 Retrospective Results

We compare the different retrospective models by examining the mean Ricker-a estimate over brood years 1978-1994, the magnitude of year-to-year variation in productivity at low spawner sizes due to the error term ($\epsilon_{i,t}$), and the year effects, m_t (these terms are from Equation [1] in Section 3.4.2). The precision of the m_t estimates and the year-to-year variation in the m_t estimates affect the precision of the treatment effect (Δm). As these sources of year-to-year variation increase, the precision of the estimate of the treatment effect decreases. In the case of the delta-style model, estimates of mean Ricker-a were for 1978-1990, because SR data for brood years 1991-1994 were unavailable for the six lower river reference stocks (Table 4-1).

Table 4-1: Comparison of retrospective models. Variances and averages are calculated over BY78-94 for the alpha-style models and BY78-90 for the delta-style model.

	Delta-style (52-90)	Alpha-style (57-94)	Alpha-style (78-94)
sigma ²	0.350	0.347	0.435
variance of m_t series 78-	1.110	1.044	0.985
variance of μ_t series 78-90*	0.629	#N/A	#N/A
variance of δ_t series 78-90*	0.221	#N/A	#N/A
mean δ_t 78-90*	-0.157	#N/A	#N/A
mean Snake a 78-	0.770	0.508	0.817
SE mean Snake a 78-	0.080	0.067	0.142

* delta-style model only

Each of the models explored give similar estimates of the variance in the m_t series. Notice that in the case of the delta-style model, m_t is partitioned into a common year effect (δ_t), and differences between Snake and lower Columbia stocks (μ_t). The variance of the series of post-1977 m_t 's is approximately equal to 1.0, and an analysis of the alpha-style (1957-1994) model showed that this variance is tightly estimated (SE ~0.10) (time series of m_t with standard errors are graphed in Figure 4-1). If one is willing to use the lower Columbia stocks as reference populations, there appears to be an advantage to defining the treatment effect in terms of changes in μ_t instead of m_t . This occurs because the variance of the μ_t series is 0.63 compared to variance of 1.04 for the m_t series estimated from the Alpha-style 1957-1994 model (Table 4-1).

The parameter estimates, standard errors, and t values are reported for each of the Alpha-style 1975-1994 model parameters (Table 4-2). Notice the tight individual estimates of the m_t parameters (SE~0.22) relative to the m_t deviation of the m_t series over time (SD~1.02). This means that the year-to-year variation in the m_t estimates, not the variance of the estimates themselves, will be most important in determining the precision of the estimate of experimental effects. Therefore, to increase the precision of the experimental response, it is fruitless to try to increase the precision of the year effects. *Only by controlling for the year-to-year variation, by designating treatment and control groups in the same year, will it be possible to increase substantially the precision of the treatment effect estimate.*

As a rule of thumb, the standard error of the experimental effect can be estimated with the formula

$$\text{var}(\Delta m \text{ estimate}) = \text{var}(m_t \text{ series (1978-1994)}) * (1/n_1 + 1/n_2) \quad [8]$$

where n_1 is the number of control years and n_2 is the number of treatment years. In the case where experiment is conducted for 10 years with treatments applied every other year starting in 2001 (5 treatment years and 28 control years), $\text{var}(\Delta m \text{ estimate}) = 1.044 * (1/28 + 1/5) = 0.246$, yielding as standard error of $\sqrt{0.246} = 0.496$. This result is in close agreement with the standard error estimate from simulations (see Table 4-23).

Table 4-2: Retrospective Results for Alpha-type Model (1957-1994).

parameter	Value	Std. Error	t value	Significance (* = sig. Diff. from 0)
Imnaha_a	1.128724	0.178105	6.337423	*
Minam_a	1.417003	0.144168	9.828823	*
Bear_a	1.216614	0.168269	7.230178	*
Marsh_a	1.029152	0.165337	6.224566	*
Sulphur_a	1.349719	0.167872	8.040178	*
Poverty_a	1.049908	0.149919	7.003157	*
Johnson_a	1.168292	0.172194	6.784729	*
m_1957	1.469921	0.278896	5.2705	*
m_1958	1.760856	0.22056	7.983558	*
m_1959	1.491049	0.223767	6.663403	*
m_1960	1.523787	0.234685	6.492916	*
m_1961	1.085287	0.226321	4.795342	*
m_1962	1.065956	0.228643	4.662092	*
m_1963	0.617317	0.225967	2.731895	*
m_1964	0.79786	0.230371	3.463368	*
m_1965	1.303694	0.220816	5.903986	*
m_1966	0.665527	0.227169	2.929659	*
m_1967	1.064311	0.228882	4.650047	*
m_1968	1.285881	0.225242	5.708891	*
m_1969	0.421099	0.224847	1.872819	
m_1970	0.431945	0.22188	1.946753	
m_1971	-0.611335	0.220663	-2.770443	*
m_1972	-1.155178	0.222066	-5.201956	*
m_1973	0.629441	0.230806	2.727143	*
m_1974	-0.909948	0.220815	-4.120864	*
m_1975	-2.056323	0.220749	-9.315225	*
m_1976	-0.826054	0.224378	-3.681524	*
m_1977	-0.531363	0.223419	-2.378329	*
m_1978	-0.853446	0.221915	-3.845825	*
m_1979	-0.622226	0.227292	-2.73756	*
m_1980	0.803279	0.22921	3.50455	*
m_1981	0.470578	0.226333	2.079142	*

parameter	Value	Std. Error	t value	Significance (* = sig. Diff. from 0)
m_1982	0.435438	0.226566	1.921901	
m_1983	1.194295	0.225592	5.294042	*
m_1984	-0.325439	0.243843	-1.334628	
m_1985	-0.770791	0.222424	-3.465413	*
m_1986	-0.261357	0.222637	-1.173911	
m_1987	-1.254864	0.22195	-5.653815	*
m_1988	-0.490098	0.22101	-2.217535	*
m_1989	-1.168014	0.226748	-5.151145	*
m_1990	-2.85249	0.22402	-12.73318	*
m_1991	-2.372832	0.224143	-10.58623	*
m_1992	-0.382679	0.224969	-1.701032	
m_1993	-0.505348	0.220159	-2.29538	*
Imnaha_b	-0.00073	0.000165	-4.41854	*
Minam_b	-0.001458	0.000241	-6.044978	*
Bear_b	-0.000646	0.000172	-3.747718	*
Marsh_b	-0.000969	0.000331	-2.922966	*
Sulphur_b	-0.002013	0.000453	-4.441713	*
Poverty_b	-0.000734	0.000148	-4.973736	*
Johnson_b	-0.002179	0.000466	-4.677157	*
sigma^2	0.347231			

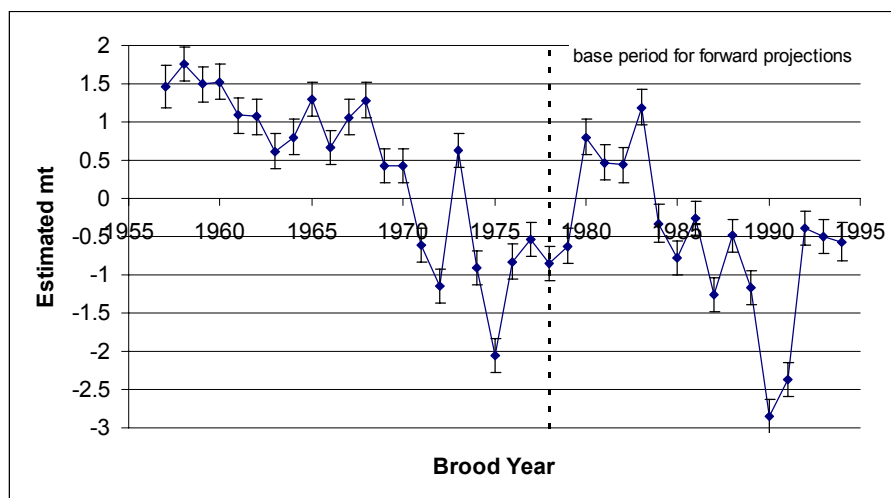


Figure 4-1: Time series of estimated m_t values, 1957-1994. Error bars are plus/minus standard error of each annual m_t estimate, from Table 4-2.

4.1.2 Population Viability Analysis

Using the Alpha-style (1957-1994) model, population viability analyses were performed using two different measures of population performance: - the CRI-type quasi-extinction measures as defined in the draft A-Fish appendix (less than a single spawner in a single year), and the probability of exceeding the 1995 BiOp recovery and survival escapement targets. We did not utilize the Alpha-style (1978-1994) version of the life cycle model because the lack of precision in the Ricker-b estimates led to unrealistic population projections. The Delta-style model was not employed because data for lower Columbia stocks were lacking from 1991-1994.

As a starting point, we defined a base case that used the relatively poor conditions of 1978-1994. To do this, we drew future m values at random from 1 million samples of retrospective (1978-1994) values of m_t (Table 4-3). We found that, under current conditions, there is a large probability that the Marsh Creek or Sulphur Creek stock will fall below the quasi-extinction threshold within the short time horizon of 10 years. Of course extinction probabilities are even larger over the 100-year horizon. The lower 95% confidence bound for the 10-year quasi-extinction probabilities are 0.527 and 0.335 for Marsh and Sulphur Creek stocks, respectively. Using the 100-year quasi-extinction metric, the lower confidence bounds are 0.659 and 0.625 for Marsh and Sulphur Creek stocks, respectively. Furthermore, none of the recovery or survival standards are met (Table 4-3) (the recovery standard is met when the probability of exceeding the recovery threshold is 0.5 or greater; the survival standard is met when the probability of exceeding the survival threshold is 0.7 or greater).

Table 4-3: Alpha-style (1957-1994) prospective results for the base case (1978-1994 conditions). 6th best stock for the survival and recovery standards are in **bold**. Results are based on 1 million samples.

crit ^{er} ion	stock	probability	lower c.l.	upper c.l.	SE
10-year quasi-extinction (prob. of <1 spawner in any of 10 years)	Imnaha	0.000	0.000	0.001	0.000
	Minam	0.001	0.000	0.004	0.001
	Bear	0.002	0.000	0.008	0.002
	Marsh	0.655	0.527	0.784	0.065
	Sulphur	0.444	0.335	0.576	0.061
	Poverty	0.000	0.000	0.002	0.001
	Johnson	0.003	0.000	0.011	0.003
100-year quasi-extinction (prob. of <1 spawner in any of 100 years)	Imnaha	0.058	0.000	0.415	0.111
	Minam	0.072	0.003	0.341	0.090
	Bear	0.083	0.000	0.429	0.125
	Marsh	0.874	0.659	0.997	0.093
	Sulphur	0.807	0.625	0.972	0.094
	Poverty	0.164	0.004	0.690	0.180
	Johnson	0.156	0.007	0.608	0.169
24-year survival	Imnaha	0.408	0.224	0.591	0.095
	Minam	0.479	0.328	0.626	0.075
	Bear	0.285	0.157	0.443	0.078
	Marsh	0.151	0.070	0.255	0.047
	Sulphur	0.172	0.081	0.279	0.051
	Poverty	0.297	0.165	0.448	0.073
	Johnson	0.283	0.160	0.426	0.068

crit ^{er} ion	stock	probability	lower c.l.	upper c.l.	SE
100-year survival	Imnaha	0.377	0.127	0.612	0.128
	Minam	0.455	0.247	0.637	0.098
	Bear	0.336	0.123	0.554	0.119
	Marsh	0.161	0.029	0.365	0.091
	Sulphur	0.224	0.072	0.380	0.081
	Poverty	0.251	0.081	0.456	0.100
	Johnson	0.274	0.108	0.455	0.090
24-year recovery	Imnaha	0.019	0.000	0.079	0.022
	Minam	0.022	0.001	0.079	0.021
	Bear	0.008	0.000	0.039	0.011
	Marsh	0.001	0.000	0.007	0.002
	Sulphur	0.005	0.000	0.022	0.007
	Poverty	0.006	0.000	0.029	0.009
	Johnson	0.019	0.000	0.072	0.019
48-year recovery	Imnaha	0.018	0.000	0.084	0.024
	Minam	0.021	0.001	0.084	0.022
	Bear	0.012	0.000	0.054	0.016
	Marsh	0.003	0.000	0.018	0.006
	Sulphur	0.008	0.000	0.039	0.011
	Poverty	0.005	0.000	0.027	0.009
	Johnson	0.019	0.000	0.080	0.021

4.1.3 Sensitivity Analysis

We performed sensitivity analysis on the Alpha-style (1957-1994) model using constant Δm values (i.e., applied in each year of the simulation period) to determine how large a boost to the Ricker-a's would be necessary for the Snake stocks to meet the jeopardy standards, and to serve as a surrogate for other actions by mapping their hypothesized Δm values to these generic results. We generated two sets of results, using two possible starting years for determining quasi-extinction and 1995 BiOp survival and recovery measures. One possibility is to start the determinations in simulated year 2000 (e.g. the 24-year survival probability would be determined over simulated years 2000-2023), the first year for which we do not have spawner estimates. Another possibility is to start in 1996 (e.g. the 24-year survival probability would be determined over simulated years 1996-2019), the year that was used in previous PATH analyses. If 1996 is used as the starting year, the first four years (1996-1999) would use actual spawner estimates for those years, rather than simulated numbers. All of the rest of the results in this report assume a starting year of 2000.

Meeting the 24-year survival standard would require a Ricker-a boost of 2.0, the 100-year survival standard, a boost of 0.8, and the 48-year recovery standard, a boost of 1.15 (Table 4-4; results for the 6th best stock are summarized in Figure 4-2). Especially troubling is the 24-year survival standard result, which implies that the current recruits per spawner at low spawning densities must be multiplied by a factor of 7.4 before the standard is met. In fact, increasing the Ricker-a values by 2.0 would mean increasing the average Snake River Ricker-a to 2.51, which is greater than the average Ricker-a for brood years 1957-1970 (2.26). The poor performance of the Marsh and Sulphur Creek populations is responsible for this very large necessary increase in the Ricker-a value. These stocks have the lowest average spawner counts for 1995-1999 among the seven Snake index stocks (Table 4.5). Marsh Creek generally has the

highest probability of extinction and the least probability of meeting its recovery and survival escapement targets (Table 4-3). This is likely due to the fact that both its average Ricker-a over 1978-1994 and its initial spawners (averaged over 1995-1999) are relatively low (Table 4.6).

Note that for Sulphur Creek and Johnson Creek stocks, there is a slight increase in 100-year quasi-extinction probability when Δm is increased from 1.5 to 2.0, which at first glance appears counterintuitive. This probably occurs because when spawner numbers are sufficiently large, the rate of recruitment (recruits-per-spawner) nears zero due to overcompensation – a characteristic of the Ricker curve (Quinn and Deriso 1999). The Ricker curve practically falls to the X-axis at high enough levels of spawners, so that reproduction is completely eliminated. Notice that for Sulphur Creek and Johnson Creek, the Ricker-b's are large (Table 4-2), thus the effect of overcompensation (the slope of the predicted line of $\ln(R/S)$ vs. S) is large. As the Δm increases, recruitment becomes more variable, and the probability that overcompensation will produce zero recruits increases. Apparently, when Δm is sufficiently large, this overcompensation effect increase the probability of extinction.

Included in the sensitivity analysis is an estimate of how the jeopardy probabilities and quasi-extinction probabilities would change if climate conditions returned to their mean (as predicted by the delta-style model). We found that this change would be insufficient for the index stocks to meet any of the jeopardy standards. To do this, we increased Δm by 0.16, which is the increase in the Δm necessary to achieve mean climate conditions as predicted by the delta-style model. Specifically, it is the magnitude of the mean delta time series over brood years 1976-1990, representing the climate effect on the downriver stocks. One must make the assumption that with a return to “normal” climate conditions, that this minor increase also applies to the Snake River stocks (that there is no Snake-specific “extra mortality” due to climate).

Table 4-4: Probabilities of meeting escapement targets as Ricker-as increase. Time periods for quasi--extinction and survival/recovery determinations start in 2000. Results are based on 1000 samples.

Criterion	Stock	Delta m				
		0.00	0.16	1.00	1.50	2.00
10-year extinction	Imnaha	0.000	0.000	0.000	0.000	0.000
	Minam	0.001	0.000	0.000	0.000	0.000
	Bear	0.002	0.001	0.001	0.001	0.001
	Marsh	0.672	0.654	0.491	0.434	0.377
	Sulphur	0.465	0.431	0.299	0.254	0.218
	Poverty	0.000	0.000	0.000	0.000	0.000
	Johnson	0.008	0.005	0.001	0.001	0.001
100-year CRI extinction	Imnaha	0.053	0.016	0.000	0.000	0.000
	Minam	0.072	0.019	0.002	0.003	0.004
	Bear	0.066	0.019	0.001	0.001	0.001
	Marsh	0.880	0.807	0.563	0.514	0.462
	Sulphur	0.817	0.737	0.529	0.511	0.532
	Poverty	0.163	0.053	0.000	0.000	0.000
	Johnson	0.136	0.054	0.014	0.051	0.124
24-year survival	Imnaha	0.414	0.490	0.782	0.842	0.869
	Minam	0.475	0.544	0.785	0.835	0.852
	Bear	0.290	0.359	0.663	0.750	0.792
	Marsh	0.154	0.195	0.476	0.605	0.682

Criterion		Delta m				
		0.00	0.16	1.00	1.50	2.00
Stock	Sulphur	0.179	0.224	0.483	0.590	0.647
	Poverty	0.293	0.361	0.707	0.793	0.833
	Johnson	0.284	0.353	0.668	0.747	0.781
	Imnaha	0.378	0.510	0.888	0.934	0.951
100-year survival	Minam	0.447	0.555	0.857	0.899	0.906
	Bear	0.340	0.470	0.857	0.914	0.936
	Marsh	0.161	0.264	0.772	0.861	0.892
	Sulphur	0.226	0.315	0.678	0.761	0.792
	Poverty	0.244	0.373	0.846	0.914	0.940
	Johnson	0.273	0.386	0.785	0.852	0.869
	Imnaha	0.017	0.048	0.686	0.914	0.979
24-year recovery	Minam	0.011	0.045	0.585	0.815	0.894
	Bear	0.012	0.029	0.555	0.891	0.973
	Marsh	0.002	0.004	0.271	0.677	0.912
	Sulphur	0.002	0.015	0.375	0.648	0.782
	Poverty	0.008	0.019	0.579	0.913	0.982
	Johnson	0.013	0.038	0.682	0.869	0.918
	Imnaha	0.020	0.046	0.698	0.925	0.984
48-year recovery	Minam	0.021	0.060	0.581	0.819	0.903
	Bear	0.012	0.040	0.641	0.897	0.979
	Marsh	0.004	0.010	0.554	0.844	0.935
	Sulphur	0.009	0.028	0.449	0.695	0.801
	Poverty	0.004	0.017	0.610	0.908	0.980
	Johnson	0.009	0.041	0.670	0.877	0.910
	Imnaha	0.020	0.046	0.698	0.925	0.984

Table 4-5: Probabilities of meeting escapement targets as Ricker-as increase. Time periods for quasi--extinction and survival/recovery determinations start in 1996. Results are based on 1000 samples.

Criterion	Stock	Delta m				
		0.00	0.16	1.00	1.50	2.00
10-year extinction	Imnaha	0.000	0.000	0.000	0.000	0.000
	Minam	0.000	0.000	0.000	0.000	0.000
	Bear	0.002	0.001	0.001	0.001	0.001
	Marsh	1.000	1.000	1.000	1.000	1.000
	Sulphur	1.000	1.000	1.000	1.000	1.000
	Poverty	0.000	0.000	0.000	0.000	0.000
	Johnson	0.003	0.003	0.001	0.001	0.001
100-year CRI extinction	Imnaha	0.053	0.013	0.000	0.000	0.000
	Minam	0.069	0.018	0.002	0.003	0.004
	Bear	0.062	0.015	0.001	0.001	0.001
	Marsh	1.000	1.000	1.000	1.000	1.000
	Sulphur	1.000	1.000	1.000	1.000	1.000
	Poverty	0.157	0.051	0.000	0.000	0.000

Criterion	Stock	Delta m				
		0.00	0.16	1.00	1.50	2.00
24-year survival	Johnson	0.128	0.050	0.013	0.047	0.118
	Imnaha	0.388	0.443	0.670	0.723	0.748
	Minam	0.438	0.490	0.679	0.723	0.739
	Bear	0.274	0.324	0.553	0.631	0.671
	Marsh	0.169	0.196	0.390	0.493	0.564
	Sulphur	0.141	0.172	0.363	0.454	0.507
	Poverty	0.330	0.380	0.643	0.717	0.754
	Johnson	0.277	0.328	0.572	0.642	0.673
100-year survival	Imnaha	0.373	0.500	0.861	0.905	0.922
	Minam	0.440	0.543	0.832	0.872	0.880
	Bear	0.336	0.460	0.830	0.886	0.906
	Marsh	0.165	0.262	0.748	0.834	0.864
	Sulphur	0.216	0.301	0.649	0.729	0.758
	Poverty	0.255	0.378	0.830	0.896	0.921
	Johnson	0.273	0.380	0.762	0.826	0.844
24-year recovery	Imnaha	0.021	0.043	0.658	0.896	0.970
	Minam	0.021	0.042	0.576	0.798	0.881
	Bear	0.005	0.017	0.459	0.830	0.973
	Marsh	0.000	0.001	0.144	0.454	0.785
	Sulphur	0.002	0.008	0.284	0.601	0.791
	Poverty	0.004	0.017	0.507	0.897	0.979
	Johnson	0.014	0.046	0.623	0.883	0.926
48-year recovery	Imnaha	0.016	0.044	0.708	0.920	0.985
	Minam	0.015	0.039	0.556	0.803	0.882
	Bear	0.011	0.042	0.648	0.888	0.977
	Marsh	0.001	0.012	0.519	0.851	0.933
	Sulphur	0.006	0.023	0.433	0.693	0.777
	Poverty	0.003	0.011	0.615	0.907	0.981
	Johnson	0.018	0.042	0.652	0.863	0.914

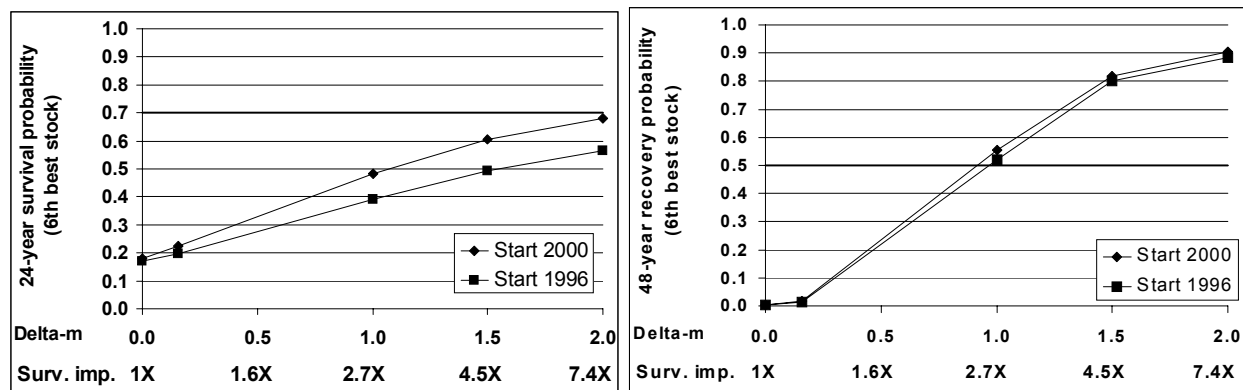


Figure 4-2: Probabilities of exceeding survival (left) and recovery (right) thresholds for the 6th best stock in each case, for various Δm values. Δm values are applied in every year of the simulation.

Table 4-6: Average Ricker-as (1978-1994) and average initial spawners (95-99)

	Ricker-a	Spawners
Imnaha	0.569	246
Minam	0.857	127
Bear	0.656	148
Marsh	0.469	58
Sulphur	0.789	42
Poverty	0.490	226
Johnson	0.608	95

4.2 Evaluation of Generic Actions

This section presents the results from power/precision analyses of a variety of “generic” management actions. The purpose of these is not to evaluate the formal statistical power of any particular action, but rather to see how long one would need to monitor some generic actions (with relatively simple schedules of Δm values) to observe an effect, measured as a change in average recruits per spawner. Results in this section use the 1957-1994 alpha-style model (with year effects drawn from 1978-1994) because:

- a) power analysis with the delta-style model is complex
- b) the differences in results for the Alpha-style and Delta-style models are not large (see Section 4.2.6), and
- c) we did not have data for lower Columbia stocks from 1991-1994.

We examined a set of six generic action/model combinations:

- 1) Alpha-style Model, 1957-94 data, 1978-94 year effects; Δm values alternate between “on” ($\Delta m = 1.0$, treated as a known constant) and “off” ($\Delta m = 0$) years (Section 4.2.1);
- 2) As above, with the “on” Δm value uniformly distributed between 0.5 and 1.5 (Section 4.2.2);
- 3) As in (1), but with measurement error reduced to zero after 1999 (Section 4.2.3);
- 4) As in (1), but with 5 years “on” alternating with 5 years “off” (Section 4.2.4);
- 5) Similar to (1), but using the delta-style model with spawner-recruit data through 1990 (run reconstructions for downriver stocks are complete only to 1990) (Section 4.2.5); and
- 6) Action effect alternates between 1.0 and 0 for 10 years, then stays at 1.0 for the duration of the simulation period.

Results suggest that differences among the 6 generic models are modest. They also suggest that at least 5 “on” (treatment) years will be needed if decision makers wish to be reasonably certain that an action is having some effect on recruitment, at least if no auxiliary information is used in the decision (see Section 4.4). By this we mean having a greater than 95% probability that the actual effect is > 0 . Recall that a Δm of 1 is equivalent to a 2.7-fold increase in spawner to recruit survival, a very substantial change. Also, because recruits return up to five years after their parents have spawned, one should add five to the

number of years in all tables, etc. for this section, since all numbers refer to brood years over which one would carry out an experiment. For example, an experiment requiring 10 brood years will have its final results 15 years after the action commences.

Although the results in this section focus on what could be learned from the experimental actions, we have also generated population projection summaries (jeopardy standards and quasi-extinction metrics). Note that these metrics assume that the experimental action will be maintained over the entire 100-year simulation. With the possible exception of the drawdown actions, this assumption is probably not realistic, because if one discovers a suite of actions that works in the sense of meeting survival and recovery requirements one likely would not continue with the original on/off experiment. Instead, one would either decide on a “final” course of action or modify the action(s) and monitoring scheme(s) based on newly acquired information. For comparison, we consider a hypothetical scenario of this type in Section 4.2.6. However, a formal analysis of this type of multi-stage decision analysis is beyond the scope of the current report. The population metrics included here may thus be viewed as a relative index of the biological consequences of experimental actions for the stocks, if the actions were continued indefinitely.

All of the population metrics are summarized in Appendix E. For comparison, and as an example of how these population metrics can be summarized along with the power analyses, we have included the 24-year survival and the 48-year recovery result for each action in this section. We report the result for the Sulphur Creek stock, because in most cases this was the 6th best stock.

4.2.1 0 in even years, 1 (exactly) in odd years.

Table 4-7 shows the series of Δm 's that we assume apply to spawner-recruit survival. As in most other cases in Section 4.2, we assume that the first treatment year is 2001, and that treatment and control years alternate in a 0/1/0/1, etc. fashion for the life of the experiment.

Table 4-7: Series of Δm 's applied to spawner-recruit survival.

Action	Year	Δm (change in surv.)
Generic 0/1	2000	0 (0X)
	2001	1 (2.7X)
	2002	0 (0X)
	2003	1 (2.7X)
		Etc.

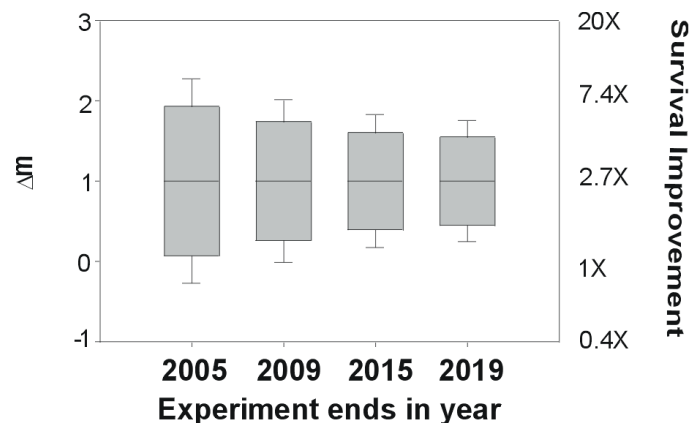
Learning

The mean and standard deviation for simulated Δm 's are shown in Table 4-8. If one employs the usual null hypothesis – namely, that one would want a <5% chance of rejecting the null hypothesis of no change in R/S survival relative to the base case (recall that the base case is the average of 1978-1994), when it has in fact changed – then 5 treatment years (ending the experiment in 2009) may be sufficient. This result may be obtained simply by multiplying the standard deviation of 0.497 by a t-score of 1.96, then subtracting the product from the mean Δm (i.e. $0.976 - (0.497 \times 1.96) < 0.05$).

Table 4-8: Results of generic action 1 ($\Delta m = 0/1$ in odd years).

Δm (change in surv.)	Year Experiment Ends	# Of Treatment Years	Estimated Δm	Std. Dev.
1 (2.7X)	2009	5	0.976	0.497
	2013	7	0.984	0.429
	2019	10	0.999	0.349
	2029	15	0.994	0.287

However, if decision makers want more precise estimates of Δm , or are interested in whether or not Δm exceeds some cut-off (e.g., 0.9), then many more years of experimentation would be needed. Figure 4-3 shows how the distribution of simulated values of Δm tighten as more years are added to the experiment. The gray box in that figure represents the range of Δm containing 90% of the estimated values. After 6 years (i.e., experiment ending in 2005), there is a 90% chance that the estimated Δm will be between 0 (this equates to a survival rate = average 1978-1994 survival rate) and +2 (7.4X base case survival rate). However, after about 20 treatment years, there will be a 90% chance that the estimated survival rate is between 1.6X ($\Delta m = 0.5$) and 4.5X ($\Delta m = 1.5$) the base case.

**Figure 4-3:** Distribution of Δm 's as the # of treatment years changes.

The distributions in Figure 4-3 can be used as the basis for making some judgements about how long this generic 0/1 on/off experiment needs to be run to detect values of Δm that reflect the risk preferences of decision-makers. Earlier in Section 3.3 we presented three example criteria for illustration:

- 1) Require the experiment to have no negative effect on survival. In this case, decision makers would want to know the probability of detecting $\Delta m \geq 0$, and how this probability changes as the experiment goes on. Looking at Figure 4-3, one can see that there is around an 80% chance of detecting $\Delta m \geq 0$ after only 1 treatment year.

- 2) Require the estimated Δm to be 80% of its hypothesized effect. In this generic action the hypothesized effect is 1.0, so one would use the distributions in Figure 4-3 to estimate how the probability of estimating $\Delta m \geq 0.8$ changes as the experiment is run.
- 3) Require the experiment to have a 0.8 probability or greater of estimating a critical value of Δm that minimizes the probability (statisticians generally like this probability to be less than 0.05) of incorrectly concluding that there is an effect, when in fact the action has no effect. For this generic action, the critical value is calculated as $1.64 * \text{the standard deviation of the distribution of estimated } \Delta m\text{'s}$ (i.e. $\Delta m^* = 1.64 * 0.497 = 0.82$ for a 10-year 1/0 on/off experiment, from Table 4-8).

Probability of detecting these effect sizes ($\Delta m = 0, 0.8$ of true, Δm^*) over time are shown in Figure 4-4. Decision-makers can use this graph to decide how long this experiment should run to achieve a desired level of certainty in detecting these “critical” effect levels. For example, if decision-makers want to be 95% confident that this action is at least doing no harm (i.e., has a 95% probability that Δm is at least 0), one would need to run the experiment for at least 6 years. Or, applying the standard statistical criteria, one would need to run the experiment for 16 years to have at least an .8 probability of detecting the critical Δm value.

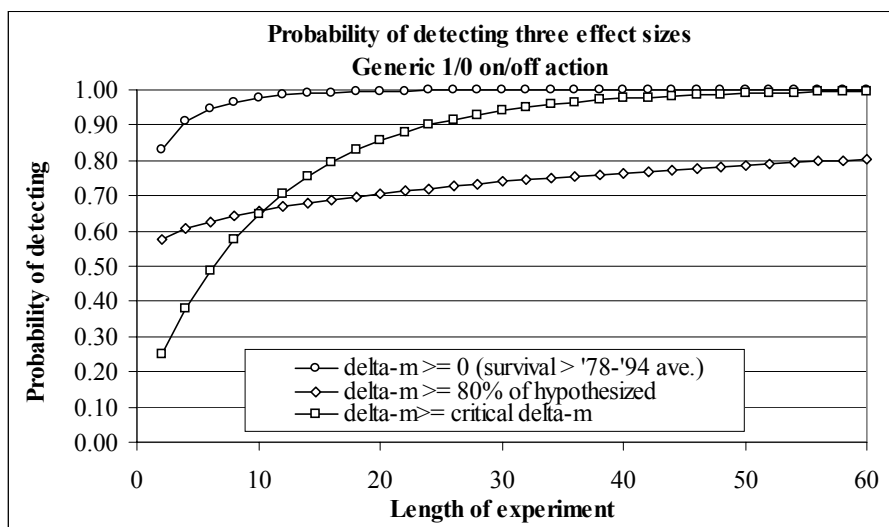


Figure 4-4: Probability of detecting three example effect sizes of Δm 's as the length of the experiment changes.

This analysis can be extended to other Δm values implemented in an on/off pattern. Figure 4-5 shows the probability of estimating critical Δm for a range of true Δm values from 0 to 5, and for lengths of experiments up to 40 years. Probabilities of estimating Δm^* increase as true Δm increases, and as the length of the experiment increases. For example, it would require 30 years of on/off experiments to get a 0.8 probability of detecting critical Δm (0.8 probability is generally desired when designing experiments) if the true value of Δm was 0.5. However, it would only take about 7 years to get 0.8 probability if the true value of Δm was 2.0.

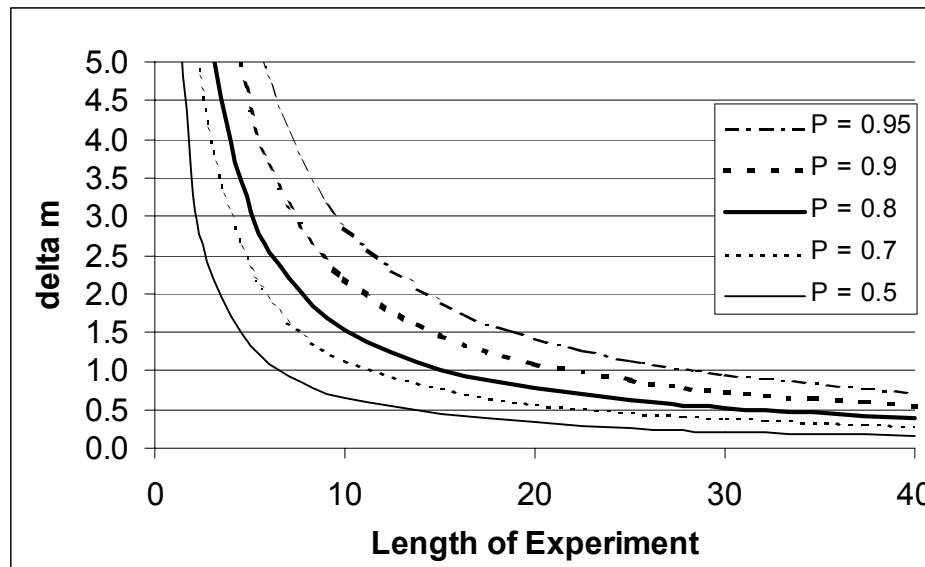


Figure 4-5. Probabilities of detecting critical Δm values for various true Δm values and lengths of experiments.

Biological

The probability of meeting the 24-year survival standard for the Sulphur Cr. stock is 0.35. Results for other stocks range from 0.33 (Marsh Creek) to 0.67 (Minam). Probability of meeting 48-year recovery standard for Sulphur is 0.15 (range 0.14 to 0.32). Again, these measures are intended only as a relative index of risk to the stock, and assumes that the experiments continue at least 48 years into the future. Consequently, the value is independent of the length of the experiment in Table 4-8.

Overall Summary

Learning and biological results are summarized in Table 4-9. This table shows the true Δm (and the corresponding change in survival, relative to the base case), several possible durations of the experiment and the corresponding # of years in which the treatment is applied, and the mean and standard deviation of the estimated Δm for each experimental duration. The summary table also shows the probability of detecting $\Delta m \geq 0$, $\Delta m \geq 80\%$ of hypothesized values, and $\Delta m \geq \Delta m^*$ (where $\Delta m^* = 1.64 * \text{std. deviation of the estimated } \Delta m$) after each experimental duration. These probabilities summarize the information from Figures 4-3 and 4-4. Finally, the summary table includes 24-year survival and 48-year recovery probabilities for Sulphur Creek stock. Results for the rest of the actions (the rest of Sections 4.2 and 4.3 of this report) are summarized in tables with a similar format as Table 4-9.

Table 4-9: Results of generic action 1 ($\Delta m = 1, 0, 1, 0$ etc.).

"True" Δm ($\Delta \text{surv.}$)	Year Exp. Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of exceeding	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
1 (2.7X)	2009	5	0.976	0.497	0.98	0.66	0.65	0.35	0.15
	2013	7	0.984	0.429	0.99	0.68	0.76		
	2019	10	0.999	0.349	1.00	0.72	0.89		
	2029	15	0.994	0.287	1.00	0.76	0.97		

4.2.2 Allow Δm to vary around a mean value of 1

One obvious limitation of the preceding action is that, even with a carefully designed experiment, the “true” Δm may vary among treatment years. For example, the effects of transportation may depend in part on in-river flows or ocean conditions. If that were true, then the effects of modifying transport regimes could vary over time, in ways that would not be apparent when the experiment is designed.

As a simple approximation to this, we examined a case where the *mean* Δm varies as previously (0/1/0/1 ...), but the treatment effect is drawn from a distribution with mean=one, and is uniformly distributed between 0.5 and 1.5. For any given year and sample, the Δm is the same for all seven index stocks. However, it varies across years (within a sample) and across samples. Results are shown in Table 4-10.

Table 4-10: Results of generic action 2 (allow Δm to vary around a mean value of 1).

"True" Δm ($\Delta_{\text{surv.}}$)	Year Experiment Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of exceeding	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
1+/-0.5 (1.6-4.5X)	2009	5	0.987	0.49	0.98	0.66	0.66	0.35	0.15
	2013	7	0.983	0.424	0.99	0.68	0.76		
	2019	10	0.988	0.369	1.00	0.71	0.86		
	2029	15	0.994	0.301	1.00	0.75	0.95		

The somewhat counter-intuitive result is that the power of the experiment is essentially the same as when Δm is constant at 1.0 exactly. There are two reasons for this. The first is the uniform distribution we assumed for the Δm 's: for every very small value (say, 0.501) that is drawn in the simulations, it is likely that a very large value (e.g., 1.499) will also be used. In effect, these cancel out when calculating the mean estimated mean value of the Δm for any given simulation run. The second reason is that the assumed variation in Δm is rather modest in comparison to the widely varying year effects (see Figure 4-1). As can be seen from the figure, the year effects vary from +1 to -3, which is a far greater range than what we have assumed for the Δm 's. This result is generally consistent with what we found for a similar simulation for the power of carcass/nutrient experiments (see October 1999 EM report, Section 3.6). The 24-year survival measure is unaffected when variation in Δm is included. If one assumes a much larger variation in the Δm 's, the results do change somewhat (not reported further).

4.2.3 Assume no measurement error after 1999

Previous PATH life-cycle modeling has assumed that approximately 40% of the apparent variation in spawner to recruit survival is caused by measurement error of the estimated spawners. To say this a bit differently, about 40% of the unexplained noise in R/S models may be associated not with process error but with errors in estimated spawner numbers. This assumption is based on estimated correlations between weir counts of adults and spawner estimates expanded from redd counts.

Originally, we had planned to complete a sensitivity analysis of what might happen to the accuracy of Δm estimates if this measurement error (and perhaps other possible errors) could be eliminated. Obviously, the assumption that one could *eliminate* measurement error is very strong. It is very unlikely that one could ever reduce it to zero. However, results in section 4.1 suggest that the year-to-year variation in the m_t estimates, not the variance of the estimates themselves, will be most important in determining the

precision of the estimate of experimental effects. Therefore, even if measurement error were eliminated it would be unlikely to increase the precision of the experimental response and thus the power of the experiments.

4.2.4 5 years of continuous “treatment”

Another obvious question is whether or not other experimental designs might yield more information than the alternating on/off design above. In addition, there may be some treatments that must run for several years in a row due to logistical constraints. We examined a design where one alternates treatment and control at 5-year intervals, as shown in Table 4-11.

Table 4-11: Series of Δm for 0/1 5 yrs on/off

	Year	Δm
5-year off, 5-year on	2000	0
	2001	0
	2002	0
	2003	0
	2004	0
	2005	1
	2006	1
	2007	1
	2008	1
	2009	1
	2010	0
	2011	0
	2012	0
	2013	0
	2014	0
	2015	1
	2016	1
	2017	1
	2018	1
	2019	1
		Etc.

The power analysis results are essentially identical to the 0/1/0/1 ... design, as can be seen in Table 4-12. On the one hand, one gains little additional precision by running experiments with several years in a row of treatment followed by several control years. On the other hand, if logistical or other constraints require such a scheme, little information would be lost thereby. However, results of this type of design may be confounded if environmental effects are autocorrelated (i.e. good years tend to be followed by good years, bad years tend to be followed by bad years).

The effect of either blocking treatment years or alternating them is not entirely neutral with respect to the survival measure. Blocking the treatment years reduced this measure from 0.35 to 0.31. This difference is small, but a similar response was shown by all other stocks (see Appendix E).

Table 4-12: Results for generic action 4 (0/1 5 yrs on/off).

"True" Δm (Δ surv.)	Year Experiment Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$ of true)	Prob ($\Delta m \geq$ Δm^*)	Prob. of exceeding	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
1 (2.7X)	2009	5	0.972	0.477	0.98	0.66	0.68	0.31	0.14
	2019	10	0.995	0.341	1.00	0.72	0.90		

4.2.5 Use of Delta-style model

Power analysis with the delta-style model closely parallels that with the alpha-style model. However, instead of comparing years effects (m_t 's) in treatment and control years, one compares μ_t 's (see Section 3.4.2 for a description of the delta-style variation of the model). The results of one sensitivity are shown in Table 4-13.

Table 4-13: Results for generic action 5 (Delta-style model; 0/1 on/off).

"True" Δm (Δ surv.)	Year Exp. Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of exceeding	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
1 (2.7X)	2009	5	0.9902	0.42	0.99	0.68	0.77	0.49	0.36

The use of the downriver stocks as a control, combined with the additional "structure" that is part the delta-style model (see Section 3.4.2) results in a modest reduction in the standard deviation of the effect size (from 0.49, Table 4-9) to about 0.42, while increasing the 24-year survival measure for Sulphur stock from 0.35 (Table 4-9) to 0.49. All stocks but one showed a similar increase in the survival measure with the Delta-style model. Note that for the delta-style model to work properly, spawner-recruit survival for the downstream stocks as a group must not change systematically over the life of the experiment. Because of the complexity of power analysis with the delta-style model, the similar results, and the lack of updated run reconstruction data for downstream stocks, we carry out power analyses of specific actions (Section 4.3) only using the 1957-1994 alpha-style model, with year effects drawn from 1978-1994.

4.2.6 0/1 experiment for 10 years, then 1 thereafter

As noted in Section 4.2, if managers some years down the road believe that they have found a suite of actions that show strong benefits for listed stocks, they would probably discontinue the experimental mode of operation, and put those actions into operation full-time. In contrast, most of the results reported here assume that the experiments will continue for decades.

As a point of comparison, we examine 24-year survival, 48-year recovery, and 100-year quasi-extinction probabilities for three different actions. The first is the generic 0/1 "forever," described in Section 4.2.1. The second is the most optimistic drawdown action, a three-year delay followed by a permanent Δm of 1.6 (see Section 4.3.6). The third is a 0/1 experiment, with 2001 being the first treatment year, and making the action (whatever it may be) permanent in 2009. The Δm series for the third action is therefore identical to the 0/1 forever experiment (Section 4.2.1) for the first 10 years, but it then stays constant at one from 2010 onward.

The results are shown in Table 4-14. Both the drawdown and the third action are considerably more optimistic than the 0/1 forever action. On the other hand, the differences between drawdown and the third action, where the “1” goes on permanently in 2009, are for the most part fairly modest, with drawdown having a slightly higher chance of meeting the PATH criteria.

Table 4-14: Results for generic action #6 (0/1 for 10 years, then 1)

	Stock	Generic #1: 0/1 (forever)	Drawdown $\Delta m=1.6$	Generic #6: 0/1 (starting 2001) to 2009, then 1 thereafter
24-Year Survival	Imnaha	0.65	0.73	0.70
	Minam	0.67	0.74	0.72
	Bear	0.52	0.62	0.57
	Marsh	0.33	0.47	0.39
	Sulphur	0.35	0.47	0.41
	Poverty	0.54	0.66	0.61
	Johnson	0.52	0.63	0.58
48-Year Recovery	Imnaha	0.32	0.95	0.70
	Minam	0.25	0.84	0.58
	Bear	0.23	0.92	0.64
	Marsh	0.14	0.87	0.53
	Sulphur	0.15	0.72	0.45
	Poverty	0.20	0.94	0.61
	Johnson	0.29	0.89	0.67
100-Year Extinction	Imnaha	0.00	0.00	0.00
	Minam	0.00	0.00	0.00
	Bear	0.00	0.00	0.00
	Marsh	0.73	0.71	0.71
	Sulphur	0.65	0.65	0.62
	Poverty	0.00	0.00	0.00
	Johnson	0.01	0.06	0.02

4.3 Evaluation of Experimental Actions

As noted previously, these results should be considered to be worked examples, rather than the final word on how one would design real, on-the-ground experiments. Most results are presented in terms of how precise the estimates of the treatment effects – the Δm 's – are likely to be if one uses only S/R data, and probabilities of detecting three example effect sizes.

As in the last section, we have also included the 24-year survival and 48-year recovery result for Sulphur Cr. as a relative index of risk and as an example of how these population metrics can be summarized along with the power analyses.

4.3.1 Base Case (Continue 1978-1994 conditions)

Because this action is essentially maintaining 1978-1994 conditions, $\Delta m=0$ and there is no overall effect to detect in the spawner-recruit data. However, one can calculate the tagging effort needed to detect

changes in D values, as estimated with Transport:Control ratios. Therefore, for this action the learning opportunities are discussed in terms of the number of PIT-tagged fish needed to detect various estimates of D and related issues, rather than the precision of estimates of Δm . We also show the 24-year survival and 48-year recovery results for Sulphur Creek, for comparison with the other experimental actions (these results are extracted from Table 4-15).

Required PIT-tag sample sizes for estimating D

This analysis is focussed on estimating D in a single year. If the intent is to estimate a mean D value over longer time periods, then there are other factors that must be considered. Some of these factors and their implications are discussed in the following section and in Appendix F.

The number of PIT-tagged fish required in treatment and control groups to ensure sufficient juveniles in each group would depend on the desired power of the test, the significance level of the test, the passage history groups being compared, whether the test were two- or one-sided, the desired minimum detectable difference between the hypothesized and true relative values of post-Bonneville survival, and the overall smolt-to-adult return rate. In determining required sample sizes, one would want to focus on D values that are critical for distinguishing between alternative actions. However, identifying these critical values at this point is difficult because:

- analyses of D are ongoing
- critical D values include the historical estimates, and no amount of future information is going to tell us what D was in the past
- we have not yet updated the modeling results with the recent (1996-1999) spawner-recruit data.

For purposes of illustrating required sample sizes, we have used hypothesized values of 0.35 and 0.65. Required sample sizes for these hypothesized (one-sided) D values, various true D values (generically, ratios of SARs), and expected return rates are given in Table 4.3.1-1.

Table 4-15: Number of PIT-tagged fish required in treatment and control groups in each year to ensure sufficient adult returns in each group, assuming 50% survival from head of Lower Granite Reservoir to Bonneville Dam tailrace for control fish. Test is one-sided, significance level is $\alpha = 0.05$, and power is $(1-\beta) = 0.80$.

Null Hypothesis	True D value	Expected LGR-to-LGR SAR for transported (treatment) group (%)					
		0.25	0.50	0.75	1.00	1.50	2.00
$D_0 \leq 0.35$	0.40	T: 277,600 C: 222,080	T: 138,800 C: 111,040	T: 92,534 C: 74,027	T: 69,400 C: 55,520	T: 46,267 C: 37,014	T: 34,700 C: 27,760
	0.50	T: 39,200 C: 39,200	T: 19,600 C: 19,600	T: 13,067 C: 13,067	T: 9,800 C: 9,800	T: 6,534 C: 6,534	T: 4,900 C: 4,900
	0.60	T: 17,200 C: 20,640	T: 8,600 C: 10,320	T: 5,734 C: 6,880	T: 4,300 C: 5,160	T: 2,867 C: 3,440	T: 2,150 C: 2,580
	0.70	T: 10,400 C: 14,560	T: 5,200 C: 7,280	T: 3,467 C: 4,854	T: 2,600 C: 3,640	T: 1,734 C: 2,427	T: 1,300 C: 1,820
	0.80	T: 7,600 C: 12,160	T: 3,800 C: 6,080	T: 2,534 C: 4,054	T: 1,900 C: 3,040	T: 1,267 C: 2,027	T: 950 C: 1,520
	0.90	T: 5,600 C: 10,080	T: 2,800 C: 5,040	T: 1,867 C: 3,360	T: 1,400 C: 2,520	T: 934 C: 1,680	T: 700 C: 1,260
	1.00	T: 4,800 C: 9,600	T: 2,400 C: 4,800	T: 1,600 C: 3,200	T: 1,200 C: 2,400	T: 800 C: 1,600	T: 600 C: 1,200

		Expected LGR-to-LGR SAR for transported (treatment) group (%)					
$D_0 \leq 0.65$	0.70	T: 900,800 C: 1,261,120	T: 450,400 C: 630,560	T: 300,267 C: 420,374	T: 225,200 C: 315,280	T: 150,134 C: 210,187	T: 112,600 C: 157,640
		T: 114,800 C: 183,680	T: 57,400 C: 91,840	T: 38,267 C: 61,267	T: 28,700 C: 45,920	T: 19,134 C: 30,614	T: 14,350 C: 22,960
	0.90	T: 46,800 C: 84,240	T: 23,400 C: 42,120	T: 15,600 C: 28,080	T: 11,700 C: 21,060	T: 7,800 C: 14,040	T: 5,850 C: 10,530
		T: 26,800 C: 53,600	T: 13,400 C: 26,800	T: 8,934 C: 17,867	T: 6,700 C: 13,400	T: 4,467 C: 8,934	T: 3,350 C: 6,700

The required total number of PIT-tagged fish released at or above Lower Granite Dam to achieve the numbers required in Table 4-15 will depend on how treatment and control groups are constructed. For example, in recent years around 10% to 15% of in-river fish have migrated undetected at Snake River dams and McNary Dam. Thus, if the control group for a particular test were to be made up of only never-detected fish, the total release for control group would be 8 to 10 times the “C” indicated in Table 4-15. Alternatively, the never-detected group could be increased by modifying downstream dams or their operations; operating in primary bypass mode or having all guidance screens removed.

Considering projections of potentially greater adult return rates in the next few years, another 5 years of marking large numbers of juvenile fish will provide information for answering some broad-scale questions. For example, if the mean annual value of D is actually 0.8, another 5 years of data will very likely allow us to rule out the value 0.35. To distinguish between mean values of 0.7 and 0.8, however, would take much longer. The following section and Appendix F present some estimates of how long it would take to make such finer-scale distinctions with varying degrees of confidence.

Effect of assuming project-specific D 's and including intra-annual variance on ability to estimate D

The analysis to derive the number of PIT-tagged smolts needed to test particular hypotheses about ‘ D ’ with a desired amount of power (summarized above and described fully in Appendix A.1) is designed for a one-project experiment (LGR), and assumes that only sampling error affects the ability to estimate a relevant D . In currently available PIT-tag data comparing SARs of transported and non-transported smolts, a significant number of fish were transported at four projects (LGR, LGS, LMN) in 1994 and three projects (LGR, LGS, LMN) in 1995 and 1996. These data suggest that:

- D differs depending on the project from which smolts are transported. Therefore, an estimate of D for LGR only may not be a good estimate of the D that the aggregate Snake River population experiences; and
- inter-annual variance in D at a given project, due to a combination of sampling and process error, may be quite large (e.g., see Bouwes et al. 1999), particularly for individual spawning stocks where spawning escapements have been extremely low in recent years. This variance may affect the length of time and/or number of PIT-tagged fish needed to reliably measure D .

We conducted analyses of recent PIT-tag data to address the effects of each of these on the results. The complete analyses and results are provided in Appendix F; here we present only a summary of the major conclusions:

- An estimate of LGR D is a reasonable approximation to overall D under the transportation scenarios analyzed, given the findings so far that LGR D and LGS D are very similar. LGS D is important because the proportion destined to be transported at LGS ranges from about 16% to 29%, depending on scenario and FGE assumption. Assumptions about D at LMN and MCN have

little effect, because the maximum average contribution of LMN is about 10%, and from MCN 5% or less (0% under A1).

- ii) Under status quo scenarios, it will likely take many more years to determine with high or even moderate confidence whether the true future D value will be sufficient to give the Snake River stocks an acceptably high probability of survival and recovery than it would take to simply determine whether D was closer to 0.35 or 0.65 (Figure 4-6). A high D value alone would not necessarily indicate that there is a high chance of survival and recovery under transportation-based options; see Bouwes et al. (1999) for other necessary assumptions.

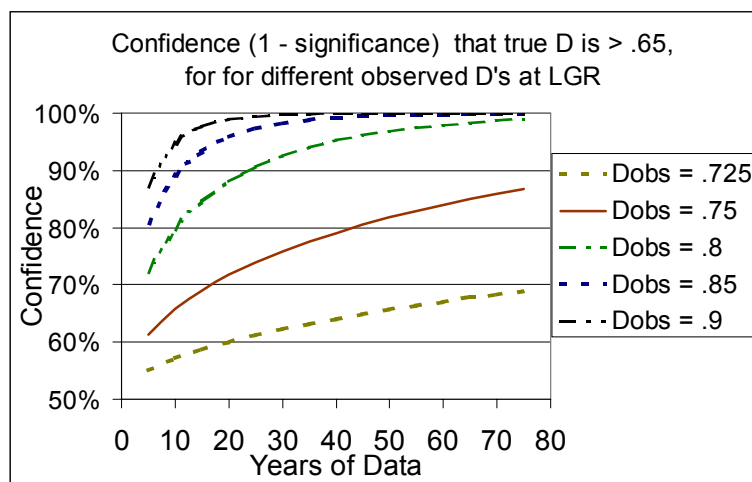


Figure 4-6: Confidence level that true D is > 0.65 at LGR for different future observed geometric means of D for a time series of given length.

Biological

This action constitutes a base case scenario, where 1978-1994 operations and conditions are assumed to continue into the future. This is equivalent to the base case defined and analyzed in Section 4.1.2, and results for all performance measures and stocks are presented in Table 4-3. Under these assumptions, the 24-year survival measure for Sulphur Cr. stock was 0.17; the 48-year recovery measure was 0.008. Both of these values are below the standard of 0.7 (survival) of 0.5 (recovery).

4.3.2 Modify transportation, measure changes in SARs

The increase in SAR for both modify transport actions (delay arrival of smolts in estuary; separate wild/hatchery smolts in barges) is assumed to be approximately 1.2-fold. Assuming that this translates directly into a Δm of 0.2 ($\approx \ln(1.2)$) for the seven index stocks, one can calculate how long it would take to estimate this reliably using stock-recruit data alone. By this we mean that if the mean effect size is 0.2, one would want a standard error of about 0.1 to be able to reject the null hypothesis that the effect size is zero with probability < 0.05 .

One can calculate an approximate answer using the equation from Section 4.1:

$$\text{Var}(\Delta m \text{ estimate}) = \text{var}(m_i \text{ series}) * (1/n_1 + 1/n_2) \quad [9]$$

where n_1 is the number of control years and n_2 is the number of treatment years. In this case, the variance is 0.1^2 , or 0.01. If n_1 and n_2 are equal, and the variance in the m_t series is 1.044, as reported in Section 4.1, then:

$$0.01 = 1.044 * (1/n_1 + 1/n_2) \quad [10]$$

so n_1 and n_2 must be approximately 200: the experiment would need 200 control years and 200 treatment years (400 years total) to detect the effect reliably. This seems to us to be well beyond all but the longest of planning horizons. It strongly suggests using PIT-tag SAR's directly to measure the action effects. The following section reports required PIT-tag sample sizes for estimating effects of separating hatchery steelhead and wild chinook in barges. Effects are measured in terms of relative SAR values, which in turn could be used to estimate incremental changes in D . The difference in survival is represented by a "survival ratio", which is the ratio of the SAR of separated fish : SAR of non-separated fish.

Required PIT-tag sample sizes for estimating D

Sample size requirements were determined using methods similar to those described in Section 3.1.1. The number of returning adults needed in each group (treatment and control) is calculated

$$n = \frac{2 \cdot (z_{(1-\beta)} + z_{\alpha})^2}{(\ln R^* - \ln R_0)} \quad [11]$$

where R^* is the "true" ratio of SARs (SAR_T/SAR_C), R_0 is the ratio hypothesized under the null hypothesis, $(1 - \beta)$ is the power, and α is the significance level. Once the n is determined, SARs must be assumed to determine the number of smolts to tag for the treatment and control groups. Table 4-16 provides yearly sample sizes for treatment and control groups under various assumptions. See Appendix F for discussion of factors affecting estimation of D over multiple years.

Table 4-16: Numbers of PIT-tagged fish required yearly in treatment and control groups to detect hypothesized levels of effects of the treatment under various assumed SARs (for control fish), and hypothesized and true levels of the effect. The control group is fish transported under current operations. The treatment group is wild spring/summer chinook transported separate from steelhead. The ratio is of the SAR of the treatment groups to the SAR of the control groups. The significance level is $\alpha = 0.05$, and the power is $(1-\beta) = 0.80$.

Hypothesized Ratio = 1.2								
True Ratio	Adults needed		Expected LGR-LGR SAR for control					
			0.25	0.5	0.75	1	1.5	2
1.25	7421	cont.:	2,968,400	1,484,200	989,467	742,100	494,733	371,050
		treat.:	2,374,720	1,187,360	791,573	593,680	395,787	296,840
1.3	1930	cont.:	772,000	386,000	257,333	193,000	128,667	96,500
		treat.:	593,846	296,923	197,949	148,462	98,974	74,231
1.35	892	cont.:	356,800	178,400	118,933	89,200	59,467	44,600
		treat.:	264,296	132,148	88,099	66,074	44,049	33,037
1.4	521	cont.:	208,400	104,200	69,467	52,100	34,733	26,050
		treat.:	148,857	74,429	49,619	37,214	24,810	18,607
1.5	249	cont.:	99,600	49,800	33,200	24,900	16,600	12,450
		treat.:	66,400	33,200	22,133	16,600	11,067	8,300
1.6	150	cont.:	60,000	30,000	20,000	15,000	10,000	7,500

		treat.:	37,500	18,750	12,500	9,375	6,250	4,688
1.8	76	cont.:	30,400	15,200	10,133	7,600	5,067	3,800
		treat.:	16,889	8,444	5,630	4,222	2,815	2,111
2	48	cont.:	19,200	9,600	6,400	4,800	3,200	2,400
		treat.:	9,600	4,800	3,200	2,400	1,600	1,200
Hypothesized Ratio = 1.0								
True Ratio	Adults needed		Expected LGR-LGR SAR for control					
			0.25	0.5	0.75	1	1.5	2
1.1	1362	cont.:	544800	272400	181600	136,200	90,800	68,100
		treat.:	495273	247636	165091	123818	82,545	61,909
1.2	372	cont.:	148800	74400	49600	37200	24,800	18,600
		treat.:	124000	62000	41333	31000	20,667	15,500
1.3	180	cont.:	72000	36000	24000	18000	12,000	9,000
		treat.:	55385	27692	18462	13846	9,231	6,923
1.5	76	cont.:	30400	15200	10133	7600	5,067	3,800
		treat.:	20267	10133	6756	5067	3,378	2,533

Biological

Assuming a $\Delta m = 0.2$ in every year, probabilities of exceeding survival and recovery thresholds are:

24-year survival (Sulphur) = 0.23 (Standard = 0.7)

48-year recovery (Sulphur) = 0.03 (Standard = 0.5)

4.3.3 Turn transportation on/off, measure D

The results of this analysis (only for life-cycle models, not "D" estimation) are shown in Table 4-17. The standard deviation in the Δm estimates is very similar to that for the generic on/off experiment (section 4.2.1). After only 10 years, the experiment has a > 90% chance of estimating some effect, and > 60% chance of estimating 80% of the true effect. However, the probability of detecting a statistically significant Δm value does not meet the usual statistical criterion of 0.8, even after 20 years. This action substantially reduces the chances of exceeding the survival and recovery thresholds because of the assumed decline in survival in non-transport years.

Table 4-17: Results for transportation on/off action.

"True" Δm ($\Delta_{\text{surv.}}$)	Year Exp. Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \leq 0$)	Prob ($\Delta m \leq -0.55$)	Prob ($\Delta m \leq \Delta m^*$)	Prob. of exceeding	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
-0.69 (0.5X)	2009	5	-0.712	0.496	0.92	0.61	0.40	0.10	0.00
	2013	7	-0.705	0.428	0.95	0.63	0.49		
	2019	10	-0.69	0.348	0.98	0.65	0.63		

4.3.4 Carcass introductions / stream fertilization

We examined 3 different experimental designs for supplementation, as shown in Tables 4-18 and 4-19.

Table 4-18: Control stocks the same for run of experiment Control stocks were Minam, Marsh, and Poverty; Imnaha, Bear, Sulphur, and Johnson were treatment stocks.

	Year	Δm	
		4 Treatment Stocks	3 Control Stocks
Parr-smolt survival increases from 0.25 to 0.30 (1.2-fold increase)	2000	0.2	0
	2001	0.2	0
	2002	0.2	0
	2003	0.2	0
	2004	0.2	0
	2005	0.2	0
	Etc.		
Parr-smolt survival increases from 0.25 to 0.50 (2-fold increase)	2000	0.7	0
	2001	0.7	0
	2002	0.7	0
	2003	0.7	0
	2004	0.7	0
	2005	0.7	0
	Etc.	Etc.	Etc.

Table 4-19: Alternate treatment and control stocks.

		Stock 1	Stock 2	Stock 3	Stock 4	Stock 5	Stock 6	Stock 7
Parr-smolt survival increases from 0.25 to 0.50 (2-fold increase), alternate treatment and control stocks	2000	0.7	0	0.7	0	0.7	0	0.7
	2001	0	0.7	0	0.7	0	0.7	0
	2002	0.7	0	0.7	0	0.7	0	0.7
	2003	0	0.7	0	0.7	0	0.7	0
	2004	0.7	0	0.7	0	0.7	0	0.7
	2005	0	0.7	0	0.7	0	0.7	0
	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.

Results are shown in Table 4-20. If there is no effect on survival, as suggested by the hypothesis based on the PATH retrospective results (Appendix G), results for this action will be identical to those for the base (no change) case in Table 4-3. Assuming a 2-fold improvement in parr-smolt survival, these experiments are virtually certain (> 0.9 probability) to estimate some positive effect and a statistically significant effect, and have > 0.8 probability of estimating 80% of the true effect. The use of treatment and control stocks helps to control for factors that cause between-year variation in all stocks (e.g. climate conditions). Designs which vary treatment / control stocks are better able to estimate 80% of the actual effect than designs that use the same treatment and control stocks for the duration of the experiments.

Implementing this action, assuming that it has a positive effect on survival, increase probabilities of exceeding survival and recovery thresholds relative to the base case. The designs that vary treatment and control stocks (we include Poverty in Table 4-20 as an example control stock) are also slightly more risk-averse in the sense that all stocks experience a modest increase in the survival measure, whereas only the treatment stocks show an increase in survival measure when treatment stocks are held constant.

Table 4-20: Results of carcass introduction/stream fertilization action.

	"True" Δm ($\Delta \text{surv.}$)	Year Exp. Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$ of true)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of exceeding S=Sulphur (treatment) P=Poverty (control)	
									24-year Survival	48-year Recovery
Carcass, no effect (i.e., base case)	0.0					0.5	0.5	0.05	S = 0.17	S = 0.008
Carcass, 1.2X increase in parr-smolt surv.	0.2	2010	10	0.186	0.171	0.88	0.59	0.99	S = 0.24 P = 0.29	S = 0.04 P = 0.0
		2020	20	0.185	0.127	0.94	0.62	0.99		
Carcass, 2X increase in parr-smolt surv., treatment-control stocks the same	0.7	2010	10	0.693	0.163	1.00	0.80	0.99	S = 0.40 P = 0.29	S = 0.30 P = 0.0
		2020	20	0.693	0.129	1.00	0.86	0.99		
Carcass, 2X increase in parr-smolt surv., vary treatment - control stocks	0.7	2010	10	0.704	0.144	1.00	0.83	0.99	S = 0.29 P = 0.45	S = 0.08 P = 0.10
		2020	20	0.706	0.104	1.00	0.91	0.99		

4.3.5 Manipulate hatchery production

Results based on the regression between m_t and hatchery releases are shown in Table 4-21. If hatcheries have had no effect on survival, as suggested by the analysis of recent passage index and SAR data (Appendix D), the results would be identical to those of the base case in Table 4-3. Based on the results in Table 4-21, one would need to run the experiment for more than 20 years to be certain (at 5%) that the smaller effect size (0.75) was greater than zero. The larger effect size could be detected reliably (i.e., < 5% chance that one would conclude the effect size was ≤ 0) within 2-4 treatment cycles. The probability of estimating a statistically significant Δm is below the 0.8 criterion for both effect sizes, even after 20 years.

If hatcheries have the upper bound effects on survival, probabilities of exceeding survival and recovery thresholds are improved from the base case but not as much as one might expect given the size of the effect (4.5-fold improvement in survival). This is because this large effect is applied only once every three years as treatments are cycled. One can use Figure 4-2 to determine what the effect on survival and recovery would be if the 4.5-fold survival improvement ($\Delta m=1.5$) were applied in each year (this would imply an action where hatchery reductions were reduced by 50% in every year of the simulation). From Figure 4-2, a Δm of 1.5 in every year would result in 0.6 probability of exceeding the survival threshold

(below the 0.7 standard), and 0.8 probability of exceeding the recovery standard (above the 0.5 standard). Such an action would have limited experimental value though because the lack of temporal contrast increases the likelihood that measured changes in survival are confounded by other factors that changed at the same time as hatchery production was reduced.

Table 4-21: Results for hatchery action.

"True" Δm ($\Delta_{\text{surv.}}$)	Year Exp. Ends	# Treatment Years	Est. Δm	Std. Dev. of est. Δm	Prob ($\Delta m \geq 0$)	Prob ($\Delta m \geq 0.8$ of true)	Prob ($\Delta m \geq \Delta m^*$)	Prob. of meeting	
								24-year Survival (Sulphur)	48-year Recovery (Sulphur)
0.0 (no effect)					0.5	0.5	0.05	0.17	0.008
0.75 (2X)	2005	2	0.72	0.76	0.84	0.58	0.26	0.41	0.26
	2011	4	0.71	0.55	0.91	0.61	0.39		
	2017	5	0.72	0.45	0.95	0.63	0.51		
	2020	7	0.72	0.42	0.96	0.64	0.56		
1.50 (4.5X)	2005	2	1.48	0.74	0.98	0.66	0.27		
	2011	4	1.49	0.52	1.00	0.72	0.42		
	2017	5	1.5	0.43	1.00	0.76	0.54		
	2020	7	1.5	0.42	1.00	0.76	0.56		

4.3.6 4-dam drawdown

Results from 4 of the possible combinations of assumptions about D, extra mortality, and length of pre-removal period (including the best and worst cases) are in Table 4-22. PVA results for all combinations are shown in Appendix "E". Note that unlike some of the previous actions, there is no cycling between treatment and control years: once the dams are removed, they remain out for the duration of the "experiment". Consequently, hypothesized survival improvements are applied in every year, in contrast to the on/off type of experimental actions where the survival improvements are applied only in treatment years. Although this leads to larger probabilities of estimating Δm effects (probabilities of estimating both a positive effect ($\Delta m \geq 0$) and a statistically significant effect ($\Delta m \geq \Delta m^*$) are both ≈ 0.8) and of exceeding survival and recovery thresholds, the lack of temporal contrast also increases the chances that any measured effects may be confounded with climate change or other changes that are coincident with dam removal. Note also that drawdown – at least if one considers it an experiment that must be monitored after dams are removed – presents some special monitoring problems, as noted in section 3.5.7.

Table 4-22: Results of drawdown actions.

	"True" Δm ($\Delta_{\text{Surv.}}$)	Year	Prob (est. $\Delta m \geq 0$)	Prob (est. $\Delta m \geq 0.8$ of true)	Prob (est. $\Delta m \geq \Delta m^*$)	Prob. of meeting	
						24-year Survival Std. = 0.7	48-year Recovery Std. = 0.5
D=0.3, 3-Year Delay BKD	1.2 (3.3X)	2010	0.99	0.69	0.79	0.41	0.56
		2015	1.00	0.75	0.95		
		2020	1.00	0.79	0.99		
D=0.3, 3-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.47	0.72
		2015	1.00	0.81	1.00		
		2020	1.00	0.86	1.00		
D=0.8, 8-Year Delay BKD	0.4 (1.5X)	2015	0.79	0.56	0.79	0.22	0.09
		2020	0.86	0.59	0.95		
		2025	0.90	0.60	0.99		
D=0.8, 8-Year Delay Hydro	1.6 (5X)	2015	1.00	0.74	0.95	0.35	0.72
		2020	1.00	0.81	1.00		
		2025	1.00	0.85	1.00		

4.3.7 Summary

Assumptions / Caveats

Overall results for all actions are summarized in Table 4-23. Again, keep in mind that these results should be viewed as worked examples, rather than as detailed experimental designs. In preparing them, we have left out many details, and assumed away a great many potential problems. Among the assumptions and caveats are the following:

- 1) We assume that the hypothetical experimental actions described in this report can actually be implemented.
- 2) We assume that an action will have some hypothesized effect, then assess how long it would take to detect that effect and how it would affect survival, recovery and quasi-extinction metrics. We have not assessed the weight of evidence in support or against the assumed magnitude of effects.
- 3) We have only looked at the effects of individual actions; combinations of actions may be more effective.
- 4) In most cases, we have only looked at how long it would take to detect effects in overall survival, from spawner-recruit data. Ideally, one would also monitor survival rates over shorter life stages to detect more immediate effects of experimental management actions.
- 5) No other, unmonitored or unknown (to the EM researchers) actions or experiments will occur concurrently. This would require substantial coordination among researchers and managers.
- 6) No climate or other natural effects will occur with the same period (on-off pattern) as the experiments. This problem is more important for experiments that cannot be turned on and off each year, such as drawdown or others that may have logistical constraints.
- 7) Spawner abundance, recruitment, aging, and other information will be gathered with at least the same intensity as at present.

Table 4-23: Summary of results for all actions.

Action	"True" Δm (Δ_{surv})	Year Exp. Ends	Prob (est. Δm ≥ 0)	Prob (est. $\Delta m \geq$ 0.8 of true)	Prob (est. $\Delta m \geq$ Δm^*)	24-year Survival Std. = 0.7	48-year Recovery Std. = 0.5
Base case (1978-1994 conditions)	0.0		0.5	0.5	0.05	0.17	0.008
Generic 1: 0/1 on/off	1 (2.7X)	2009	0.98	0.66	0.65	0.35	0.15
		2013	0.99	0.68	0.76		
		2019	1.00	0.72	0.89		
		2029	1.00	0.76	0.97		
Generic 2: Generic 1 w/ uniform dist.	1 +/- 0.5 (2.7X)	2009	0.98	0.66	0.66	0.35	0.15
		2013	0.99	0.68	0.76		
		2019	1.00	0.71	0.86		
		2029	1.00	0.75	0.95		
Generic 4: 0/1; 5 yrs on/5 yrs off	1 (2.7X)	2009	0.98	0.66	0.68	0.31	0.14
		2019	1.00	0.72	0.90		
Generic 5: Generic 1 w/Delta-style model	1 (2.7X)	2009	0.99	0.68	0.77	0.49	0.36
Generic 6: 0/1 for 10 years, then 1	1 (2.7X)					0.41	0.45
Modify Transport	0.2 (1.2X)					0.23	0.03
Transport on/off	-0.69 (0.5X)	2009	0.92	0.61	0.40	0.10	0.00
		2013	0.95	0.63	0.49		
		2019	0.98	0.65	0.63		
Carcass: No effect	0.0		0.5	0.5	0.05	Sulph. = 0.17 Pov. = 0.29	Sulph = 0.008 Pov. = 0.005
Carcass 1: 1.2X parr- smolt survival treatment stocks constant	0.2 (1.2X)	2010	0.88	0.59	0.99	Sulph = 0.24 Pov. = 0.29	Sulph. = 0.04 Pov. = 0.005
		2020	0.94	0.62	0.99		
Carcass 2: 2X parr- smolt survival treatment stocks constant	0.7 (2X)	2010	1.00	0.80	0.99	Sulph = 0.40 Pov. = 0.29	Sulph. = 0.30 Pov. = 0.005
		2020	1.00	0.86	0.99		
Carcass 3: 2X parr- smolt survival treatment stocks vary	0.7 (2X)	2010	1.00	0.83	0.99	Sulph = 0.29 Pov = 0.45	Sulph = 0.08 Pov = 0.10
		2020	1.00	0.91	0.99		
Manipulate hatchery production	0.0		0.5	0.5	0.05	0.17	0.008
	0.75 (2.1X)	2005	0.84	0.58	0.26	0.41	0.26
		2017	0.95	0.63	0.51		
		2020	0.96	0.64	0.56		
	1.50 (4.5X)	2005	0.98	0.66	0.27		
		2017	1.00	0.76	0.54		
		2020	1.00	0.76	0.56		

Action	"True" Δm (Δ_{surv})	Year Exp. Ends	Prob (est. Δm ≥ 0)	Prob (est. $\Delta m \geq$ 0.8 of true)	Prob (est. $\Delta m \geq$ Δm^*)	24-year Survival Std. = 0.7	48-year Recovery Std. = 0.5
D=0.3, 3-Year Delay BKD	1.2 (3.3X)	2010	0.99	0.69	0.79	0.41	0.56
		2015	1.00	0.75	0.95		
		2020	1.00	0.79	0.99		
D=0.3, 3-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.47	0.72
		2015	1.00	0.81	1.0		
		2020	1.00	0.86	1.0		
D=0.8, 8-Year Delay BKD	0.4 (1.5X)	2010	0.79	0.56	0.79	0.22	0.09
		2015	0.86	0.59	0.95		
		2020	0.90	0.60	0.99		
D=0.8, 8-Year Delay Hydro	1.6 (5X)	2010	1.00	0.74	0.95	0.35	0.72
		2015	1.00	0.81	1.0		
		2020	1.00	0.85	1.0		

General Conclusions

Assuming that these conditions and assumptions can be satisfied in practice, the life-cycle results above have a number of important points in common. The list below assumes that ancillary information on life-stage survival is not brought to bear on the problem. Section 4.4 considers what might happen if such information were used directly in the design and monitoring process.

Biological

1. More than a 7.5-fold improvement in life-cycle survival is needed to meet the 24-year survival standard of 0.7.
2. A 2.7-fold increase in life-cycle survival is needed to meet the 48-year recovery standard of 0.5.
3. Survival and recovery probabilities in this analysis are lower than previous PATH results because:
 - assumes poor 1978-1994 ocean conditions continue
 - assumes extra mortality here to stay
 - uses updated spawner-recruit data
4. All of the actions except the transport on/off action provide some improvement in survival and recovery probabilities relative to the base case (continue current operations). However, over the long term none of the actions meet the survival standard (0.7) for the weakest stocks. Only drawdown meets the recovery standard (0.5) if a low historical D (D=0.3) or the Hydro extra mortality hypothesis is assumed. Probabilities of exceeding survival and recovery thresholds for the transport on/off, carcass introduction (treatment and control stocks varied), and hatchery actions assume that these actions are implemented as on/off experiments for the duration of each metric's time horizon. This reduces their effects on survival and recovery probabilities but also reduces the possibility that estimated effects are confounded by coincidental changes in other parts of the system. This is probably not a realistic assumption because if an action appeared to be increasing survival it would likely be turned on permanently.

Learning

1. Most experiments have >0.8 probability of estimating some survival improvement (i.e. $\Delta m > 0$) within 5-10 years.
2. Actions that generate > 4 -fold survival improvement (i.e., some hypothesized responses to 4-dam drawdown and reductions in hatchery output) have about a 0.8 probability of estimating Δm of at least 80% of true value after 20 years. This is also true for actions that have smaller survival improvements but have spatial controls (i.e., carcass introductions/stream fertilization).
3. Actions that generate ≤ 2 -fold survival improvements with no spatial controls (i.e., transport / no transport, and some hypothesized responses to drawdown and hatchery reductions) have about a 0.6 probability of estimating Δm of at least 80% of true value after 20 years.
4. The probabilities of detecting a statistically significant Δm are low (i.e. less than the 0.8 criterion generally applied by statisticians) for all on/off experiments except for carcass introduction/stream fertilization. The use of spatial controls in that experiment improves the ability to estimate effects. These probabilities are high for drawdown because the hypothesized survival effects are large and are applied in every year, rather than in every other year as with the on/off type of experimental actions.
5. More complex designs / expanded monitoring of life-stage specific survival data is needed to improve the ability to detect effects.
6. The ability to detect effects of management actions that affect all Snake spring-summer chinook stocks simultaneously using spawner-recruit data alone is limited. This appears to us to be inherent in the data: spawner-recruit survival varies enormously from year to year, and any attempt to measure the effects of actions must cope with these highly variable survivals. In addition, these data are affected by factors outside of direct management control such as climate and ocean conditions. Supplementing spawner-recruit information with other data sources may increase our ability to detect effects; possible approaches for doing this are explored further in Section 4.4.
7. Wherever possible, within-year comparisons (e.g., treatment and control stocks for carcass introductions, treatment and control tag groups for hatchery/wild separation in barges) should be used to control for between-year variability and thus improve the ability of the action to estimate effects.
8. For status quo and modify transport options, large numbers of PIT-tagged fish may be required to detect effects on SARs, depending on assumptions about future SARs and what groups are used as controls. The largest estimates of tagged fish required may not be feasible.
9. Some tradeoffs are evident. Clearly the transport/no-transport action presents increased risk to the populations, at least given our assumptions for that action. On the other hand, varying treatment and control stocks in the carcass experiment is a win-win option – it is slightly more risk-averse in that it spreads the assumed benefits among all stocks, while also improving the precision of estimated effects.
10. We have not yet analyzed combinations of actions, which (if done carefully) may improve the ability to detect effects and meet survival and recovery standards. The SRP suggested two possible ways of combining actions: an incremental approach (individual actions are implemented one at a time and monitored for effects), and a “reverse staircase” approach (a cluster of actions are implemented all at once, then individual actions are halted one at a time). The SRP suggested that the reverse staircase

approach may be more risk-averse, and ultimately more cost-effective, than the incremental approach. Further explorations of these two general strategies are needed.

4.4 General Discussion

Given the high level of variability in the Δm series, it will take some years to detect the effects of a management action, however large those effects might be. This section is an initial exploration of some possible methods for increasing the precision of our estimates of the effects of management actions. We suggest two broad methods. The first is more complex but potentially more informative experimental designs. The second, a close corollary of the first, is an expansion and re-direction of PIT tagging efforts in support of uncovering the effects of experimental management actions. While we do not explicitly address combinations of actions, these can be analyzed in much the same way as single actions in isolation.

Two examples of more complex design come to mind. Recall that in most of the examples in Section 4.3, the entire Snake spring/summer chinook population was treated as a single experimental group in any given year. The exceptions to this rule – for Sections 4.3.2 and 4.3.4 – allow for the possibility of treating different populations differently, either in transport groups (with and without hatchery steelhead on the barges) or in the carcass/nutrient enhancement. Based on results to date, we conclude that almost any design that has multiple treatment/control groups within a calendar year will substantially enhance the power of the experiment. Only in this way can one “control” for the large, inter-annual differences in spawner-recruit survival, at least without the use of additional information from fish not directly involved in the experiments.

A second, more complex design was outlined in Section 4.2.3 – reducing measurement error in aid of increasing the power of an experiment. As noted there, this would involve additional monitoring – some of it already ongoing – beyond what was done in the past. Developing this notion completely would be a small research project in its own right, but several areas seem obvious. In addition to reducing potential errors in the enumeration and aging of spawners, one very promising area is direct estimation of upstream survival (starting at Bonneville Dam). This probably has merit in its own right, since estimates of upstream mortality for spring/summer chinook appear to be in the same ballpark as estimates of downstream survival for in-river migrants. It seems reasonable that actions to increase upstream survival are dependent in part on more complete knowledge of where and when that mortality occurs. From an experimental management viewpoint, direct measurement would help reduce potential errors in conversion-rate based estimates of mortality, as these are very complex. They require data on dam counts, harvest rates, and tributary turnoff for 5-10 stocks that leave the Columbia at or below McNary Dam. They also necessarily assume that all fish of a particular stock group (e.g., Snake spring chinook) have the same survival and harvest rate in any given return year. Given the diversity in downstream run timing, population dynamics, and other characteristics of the index stocks, it seems plausible that their upstream survivals may show some differences as well. Again, it is not likely that these measures will eliminate measurement error, but they have the potential to reduce it substantially.

Another possibility is to account for the variation in year effects – the m_t 's – by constructing (nearly) independent estimates of their values from other data sources. Figure 4-7 and Table 4-24 show time-series of year effects and SAR's from several PIT-tag based sources. The year effects (estimated in natural log units) are exponentiated for convenience of comparison. SAR's are drawn from the Sandford and Smith 1999 draft (for wild fish transported from LGR and LGS), and for both hatchery and wild fish transported from the Snake projects, weighted as:

- 1 for fish transported at Lower Granite;
- 0.5 for fish transported at Little Goose; and
- 0.25 for fish transported from Lower Monumental.

The intent of the weighting is to have the tagged, transported fish in the sample be roughly representative of the run at large, with about ½ of the run encountering each project being transported there.

Table 4-24: Time-series of year effects and SAR's from several PIT-tag based sources.

Brood Year	1957-1994 Alpha-style Model Year Effect Estimate	Std Err	Passage Year	Exp (Year Effect)	SAR - Trans Wild Chinook-Fr S&S Tables A1 and Table 2	SAR - Hatchery fish trans at LGR, LGS, LMN - Weighted	SAR - Wild fish trans at LGR, LGS, LMN - Weighted
1987	-0.545	0.228	1989	0.58		0.02	0.12
1988	0.706	0.239	1990	2.03	0.4	0	0.59
1989	0.029	0.224	1991	1.03	0.35	0.11	0.28
1990	-1.658	0.225	1992	0.19	0.08	0.04	0.14
1991	-1.177	0.225	1993	0.31	0.11	0.07	0.11
1992	0.828	0.224	1994	2.29	0.64	0.14	0.64
1993	0.710	0.233	1995	2.03	0.36	0.53	0.37
1994	0.606	0.241	1996	1.83		0.14	0.14

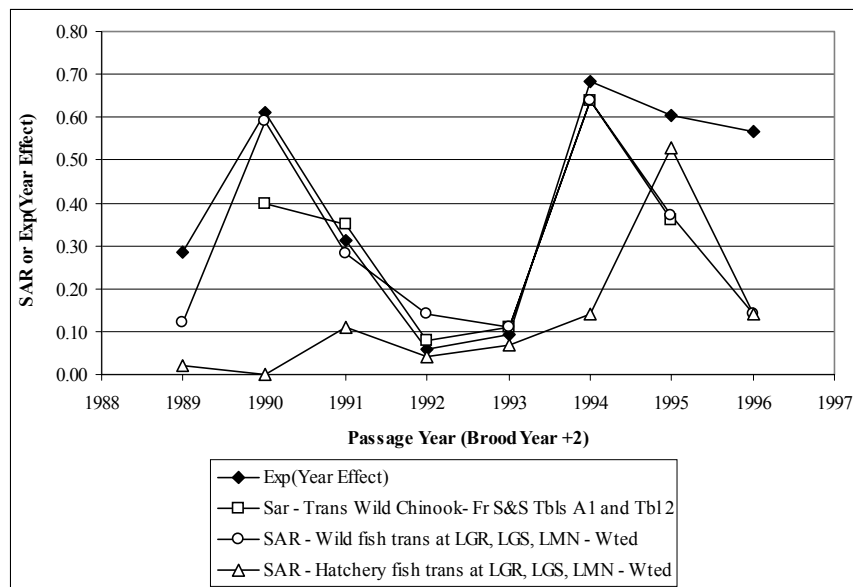


Figure 4-7: Year Effects (m_t 's) vs. PIT-tag SAR's. See text for details.

As one can see from the figure, the SAR's and year effects co-vary closely over time, with correlation coefficients as high as 0.97 between the Sandford and Smith series and the year effects. Correlations between the year effects and the other series are lower, on the order of 0.6 – 0.8, but all of the SAR series have roughly the same temporal pattern as the year effects.

While we do not want to give the impression that the SAR's *explain* the year effects, it should be possible to exploit their co-variation in experimental designs that rely, in large part, on spawner-recruit data and estimates of year effects. Without going into detail, one could design a joint estimation procedure that relies in part on the SAR estimates to condition estimates of year effects. This could reduce some of the inter-annual variation in the estimates of the year effects themselves, and hence increase the precision of the Δm estimates. It would be particularly interesting to explore this using SAR's estimated for hatchery fish in combination with management actions directed exclusively at wild stocks (e.g., nutrient enhancement or habitat improvement).

The parr-smolt survivals estimated from PIT tag data – used in the carcass enhancement experiments – provide another hint at ways to reduce random noise in the designs. They provide additional direct life-stage survival estimates that are independent of the spawner-recruit data, and may provide a means to factor out some of the very large year-to-year variability in life cycle survival.

At present, the PIT-tagging program for spring/summer chinook is focused on three major areas: arrival timing at Lower Granite (the reason for tagging wild parr in the subbasins), downstream survival of in-river migrants, and, most recently, comparing SAR's for transported fish and in-river migrants. If the program were to operate, in part, in support of experimental management activities, some additional areas of emphasis would likely be required. The overall point would be to obtain accurate estimates of life-stage survivals (e.g., egg-parr, parr-smolt, etc.) in all life stages likely to be affected by EM actions. These might include:

- 1) Subbasin parr tagging and recapture sufficient to actually enumerate parr population sizes. This would be needed if management actions were thought to influence egg-parr survival. At present, the program does not try to make estimates of the parr population sizes.
- 2) Tagging sufficient wild parr in the subbasins to obtain precise estimates of parr-smolt survival for all seven index stocks and as many other spring-summer chinook stocks as possible.
- 3) As a corollary to (2), tagging sufficient known-origin wild parr (or, less plausibly, outmigrating smolts) to obtain stock-specific in-river survival estimates and perhaps SAR's by larger groupings (e.g., Salmon Middle Fork vs. Salmon South Fork). SAR's are likely to be problematic absent improvements in Granite-to-Granite survival.
- 4) Reliable, long-term detection of adults at Bonneville, McNary, Priest Rapids, and perhaps Ice Harbor, to estimate both upstream survival and straying rates.

Items (1) and (2) would be most relevant for habitat or nutrient enhancement, while (3) and (4) would be important for actions that are thought to affect smolt-to-adult survival. It might be possible to use Clearwater stocks for a portion of the experiments in either or both areas, since they are not part of the ESU and hence may have fewer regulatory restrictions.

In evaluating both specific proposals and the EM concept in general, it is important to keep in mind what the entire experimental management program is intended to do. If we were certain what the effects of nutrient enhancement, hatchery manipulation, drawdown, etc. would be on listed stocks, there would be no need to undertake any experiments. Decision makers would simply decide what suite of actions represented a reasonable balance between biological goals – stock survival and recovery, etc. – and other considerations, and pursue the same set of actions forever. Indeed, there would be no need to monitor the results of the actions, since those would be known before the actions were initiated.

Unfortunately, given the often-ambiguous evidence associated with most management actions, there is much uncertainty about the effects of many if not most possible actions. The rationale for experimental management is that it can provide a means to resolve those uncertainties in a timely fashion, without posing unreasonable risks to listed stocks. The methodological problems – the focus of this report – can probably be resolved. If one designs and monitors a suite of EM actions carefully, one can probably measure their effects fairly precisely in a reasonable amount of time. In contrast, we think that it is unlikely that a status quo strategy – continuing current operations and waiting to see what happens – will resolve the uncertainties that PATH, CRI, the ISAB and other regional science groups have identified.

5.0 Next Steps

If further work on experimental management is undertaken after PATH ends, we recommend that effort be focussed on resolving / addressing the limitations of our preliminary analysis. Specifically, we suggest the following next steps:

1. Complete an assessment of the feasibility of implementing these experimental actions. For some actions, this will require consulting with regional management groups (e.g., hatchery managers, private and public landowners for carcass introductions/stream fertilization).
2. Assess the evidence in support / against our hypothesized effects of actions. The hypothesized values used in this report were suggested only as examples of values that might be used and approaches that could be used to develop hypotheses. Closer scrutiny of these and other hypotheses is needed. However, the hypothesized effects of most actions considered here are unlikely to be resolved without a series of well-planned experimental actions.
3. Use the model we have developed to explore alternative experimental designs and combinations of actions. There are many possible alternative designs to the ones we have used in our analyses, and many possible combinations of actions that could be explored (some of these combinations were discussed in the October 1999 Experimental Management Scoping Report). By strategically combining some of the experiments, one could test for multiple effects simultaneously.
4. Explore other monitoring to detect effects. Given the many factors that affect spawner-recruit data, and the large variability in spawner-recruit survival, the effects of actions on life-stage specific survival rates should be monitored in addition to the effects on spawner-recruit data. Such life-stage specific information may improve our ability to estimate the immediate effects of actions more precisely than the spawner-recruit data, although monitoring spawner-recruit data is still needed to assess overall survival responses.

6.0 References

- Bouwes, N., H. Schaller, P. Budy, C. Petrosky, R. Kiefer, P. Wilson, O. Langness, E. Weber, and E. Tinus.** 1999. An Analysis of Differential Delayed Mortality Experienced by Stream-type Chinook Salmon of the Snake River: Response to A-Fish Appendix. October 4, 1999
- Efron, B and R.J Tibshirani.** 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.
- Deriso, R.** 1997. Prospective Analysis of Spring Chinook of the Snake River Basin. In Marmorek, D. and C. Peters (eds.) Plan for Analyzing and Testing Hypotheses (PATH): Retrospective and Prospective Analyses of Spring /Summer Chinook Reviewed in FY 1997. Compiled and edited by ESSA Technologies Ltd., Vancouver B.C.
- Deriso, R., D. Marmorek, and I. Parnell.** 1996. Retrospective analysis of passage mortality of spring chinook of the Columbia River. In: Plan for Analyzing and Testing Hypotheses Final Report on Retrospective Analyses, Chapter 5. ESSA Technologies, Vancouver, BC, Canada.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin.** 1995. Bayesian Data Analysis. Chapman & Hall, New York.
- Petrosky, C. and H. Schaller.** 1998.
- Quinn, T.J and R.B Deriso.** 1999. Quantitative fish dynamics. Oxford University Press, New York.
- Sandford and Smith.** 1999. draft.
- Walters, C.J.** 1986. Adaptive Management of Renewable Resources. MacMillan Publishing Company, Ltd, New York.

Appendix A: Complete Descriptions of Actions

A.1 Continue Current Hydropower Operations and Estimate Post-Bonneville Dam Survival

A.1.1 Description of experimental action / research & monitoring

Rationale

Study Objective: To estimate post-Bonneville Dam survival of smolts under current operating conditions with various juvenile passage histories, with particular emphasis on comparison of transported smolts versus those that migrate in the river (D). This analysis is focussed on estimating D in a single year. If the intent is to estimate a mean D value over longer time periods, then there are other factors that must be considered. Appendix F discuss some of these factors and their implications.

Description of Hypothesis: Post-Bonneville Dam survival relates to two immediate critical issues regarding current operations: (1) differential mortality of transported smolts and in-river migrants (D); and (2) general evidence of delayed effect of hydropower system on life-cycle survival. Specifics in this section deal primarily with investigation of D , though many issues will be relevant to the general question.

With regard to D , several null hypotheses are possible. One is that post-Bonneville survival of transported fish is the same as those that migrate in the river. However, this null would be somewhat artificial, as there is consensus that post-Bonneville mortality of transported fish is at least slightly elevated. Based on previous PATH decision analysis results and recent analyses of PIT-tag data suggesting that recent and prospective D -values might be higher than previously estimated, the most useful form for the null hypothesis is one-sided:

$$H_0 : D \leq D_0 \quad [A1]$$

where D_0 is the hypothesized value of D . Thus, rejection of this hypothesis constitutes evidence that the true value of D is greater than D_0 . Several values of D_0 are worth considering, and one would want to focus on D values that are critical for distinguishing between alternative actions. For purposes of illustrating required sample sizes, we have used hypothesized values of 0.35 and 0.65.

The results of the decision analyses are sensitive to the value of D . Methods of transportation have improved, as has survival of downstream migrants (and the means to estimate it), so that estimates of D derived from earlier studies, particularly pre-1980, do not apply to the present or in the future. Direct losses in the hydropower system in the 1970s and early 1980s clearly impacted stocks; however, current estimates of downstream migrant survival are similar to survival of fish through the hydropower system prior to construction of John Day, Lower Monumental, Little Goose, and Lower Granite Dams. Transport studies conducted under a range of in-river and ocean conditions would reduce uncertainty about the efficacy of transportation.

Other investigations involving comparisons of post-Bonneville Dam survival for groups of fish with different juvenile passage histories might use comparable hypotheses. For example, the simplest test would involve the two-sided null hypothesis of equality of Bonneville-to-Lower Granite SAR for two passage history groups.

Experimental Action: Continue transport evaluation studies in the Snake River using PIT tags for both yearling chinook salmon and steelhead. Conditions for inriver migrants would be optimized by maximizing spill at downstream projects during the migration.

Spatial and Temporal Components

Large-scale juvenile tagging programs will be ongoing for the foreseeable future for various purposes, including evaluating transportation programs. Assuming five years of juvenile fish marking to evaluate transportation and downstream migrant survival (2000 through 2004), complete adult evaluation will occur by 2007. The work should encompass PIT-tagged fish from all major tributaries of the entire Columbia River Basin above Bonneville Dam. The experimental units will include some combination of streams, cohorts, stocks, and juvenile passage histories, evaluated within and between years.

A.1.2 Monitoring approach

Variables to Monitor

For each year of the study, use NMFS method to estimate post-Bonneville Dam survival for PIT-tagged smolts with each juvenile passage history. The general method has several steps, including estimation of survival where possible through reaches of the lower Snake River and lower Columbia River, extrapolation of survival estimates outside reaches where empirical estimates are possible, estimation of the number of fish that arrived at Lower Granite Dam, and estimation of the “Lower Granite-equivalent” number of PIT-tagged fish that experienced each possible passage history during juvenile migration, and calculation of Bonneville (smolt)-to-Lower Granite (adult) return rates for each group. For estimating and testing D , appropriate “treatment” and “control” groups of PIT-tagged fish must be constructed, with the objective of creating groups that are representative of transported and “in-river” fish in the run at large. Further detail on methods is available elsewhere.

PIT tags provide timing information for all detected fish. Relationships among travel time, survival, passage history, and environmental conditions (including flow, water temperature, and levels of spill) would also be evaluated. Reach survival would be estimated using the PIT-tag interrogation system now in place in the Snake and Columbia Rivers (with the planned upgrade to the new PIT-tag frequency), the PIT-tag towed array in the Columbia River estuary, and recoveries of tags from bird colonies below Bonneville Dam. If tag detections below Bonneville Dam are sufficient, survival to Bonneville Dam will be estimated directly. Otherwise, survival estimates to Bonneville Dam will be extrapolated from upstream reach estimates.

Other patterns in differential post-Bonneville Dam survival will be investigated. Differences in SAR related to number of times detected (or other patterns) may provide general evidence of delayed effects of the hydropower system. The level of detail provided by PIT-tag data may show that post-Bonneville survival depends on the time of arrival below Bonneville Dam. For example, survival of transported smolts relative to in-river smolts may vary within a single season.

Sample Sizes Required

Required sample sizes would depend on the desired power of the test, the significance level of the test, the passage history groups being compared, whether the test were two- or one-sided, the desired minimum detectable difference between the hypothesized and true relative values of post-Bonneville survival, and the overall smolt-to-adult return rate. Again, the sample sizes calculated here are for estimating D in a single year; other factors must be considered for estimating a mean D over longer time periods.

Using NMFS' methods, variance of estimated D is estimated using bootstrap methods, but can be approximated by variance of the ratio of probability estimates (Bonneville-to Lower Granite SARs for treatment and control groups) from two independent binomial distributions. The equation for variance of the ratio of two SARs is (Burnham et al. 1987):

$$Var\left(\frac{\hat{p}_1}{\hat{p}_2}\right) = \left(\frac{1}{n_1} - \frac{1}{N_1} + \frac{1}{n_2} - \frac{1}{N_2}\right) \left(\frac{\hat{p}_1}{\hat{p}_2}\right)^2 \quad [A2]$$

where N_i is the number of juveniles, n_i is the number of returning adults, and p_i is the return rate in group i . Note this ignores the fact that the number of juveniles is estimated rather than known, but as the number of adults is usually much greater than the number of juveniles, including uncertainty in the estimated juveniles adds little to the variance of the ratio. Analyses are typically done on the log-transformed scale. The variance of the log of the ratios is:

$$Var\left(\ln \frac{\hat{p}_1}{\hat{p}_2}\right) = \left(\frac{1}{n_1} - \frac{1}{N_1} + \frac{1}{n_2} - \frac{1}{N_2}\right) \quad [A3]$$

Because return rates are quite low (i.e., $N_i \gg n_i$), this expression is dominated by the inverse of counts of returning adults in the two groups. If we plan the study so that the expected number of adults in each group is equal ($n_i = n$), then the right side of the equation reduces to approximately $2/n$ (see Section 3.2.2). For example, if we hypothesize $D_0 \leq 0.35$, and we wish to have 80% power to detect a difference if the true D is 0.65 or greater with a 0.05-significance level test, then the number of adults needed in each group is (Steel and Torrie 1980):

$$n = \frac{2 \bullet (z_{.80} + z_{.95})^2}{(\ln(0.65) - \ln(0.35))^2} = 33. \quad [A4]$$

For two-sided tests at $\alpha=0.05$, $z(.95)$ is replaced by $z(.975)$. The number of PIT-tagged fish required in treatment and control groups to ensure sufficient juveniles in each group for various hypothesized (one-sided) and true D values (generically, ratios of SARs) and expected return rates are given in Table A-1.

Table A-1: Number of PIT-tagged fish required in treatment and control groups to ensure sufficient adult returns in each group, assuming 50% survival from head of Lower Granite Reservoir to Bonneville Dam tailrace for control fish. Test is one-sided, significance level is $\alpha = 0.05$, and power is $(1-\beta) = 0.80$.

Null Hypothesis	True D value	Expected LGR-to-LGR SAR for transported (treatment) group					
		0.25	0.50	0.75	1.00	1.50	2.00
$D_0 \leq 0.35$	0.40	T: 277,600 C: 222,080	T: 138,800 C: 111,040	T: 92,534 C: 74,027	T: 69,400 C: 55,520	T: 46,267 C: 37,014	T: 34,700 C: 27,760
	0.50	T: 39,200 C: 39,200	T: 19,600 C: 19,600	T: 13,067 C: 13,067	T: 9,800 C: 9,800	T: 6,534 C: 6,534	T: 4,900 C: 4,900
	0.60	T: 17,200 C: 20,640	T: 8,600 C: 10,320	T: 5,734 C: 6,880	T: 4,300 C: 5,160	T: 2,867 C: 3,440	T: 2,150 C: 2,580
	0.70	T: 10,400 C: 14,560	T: 5,200 C: 7,280	T: 3,467 C: 4,854	T: 2,600 C: 3,640	T: 1,734 C: 2,427	T: 1,300 C: 1,820
	0.80	T: 7,600 C: 12,160	T: 3,800 C: 6,080	T: 2,534 C: 4,054	T: 1,900 C: 3,040	T: 1,267 C: 2,027	T: 950 C: 1,520
	0.90	T: 5,600 C: 10,080	T: 2,800 C: 5,040	T: 1,867 C: 3,360	T: 1,400 C: 2,520	T: 934 C: 1,680	T: 700 C: 1,260
	1.00	T: 4,800 C: 9,600	T: 2,400 C: 4,800	T: 1,600 C: 3,200	T: 1,200 C: 2,400	T: 800 C: 1,600	T: 600 C: 1,200
$D_0 \leq 0.65$	0.70	T: 900,800 C: 1,261,120	T: 450,400 C: 630,560	T: 300,267 C: 420,374	T: 225,200 C: 315,280	T: 150,134 C: 210,187	T: 112,600 C: 157,640
	0.80	T: 114,800 C: 183,680	T: 57,400 C: 91,840	T: 38,267 C: 61,267	T: 28,700 C: 45,920	T: 19,134 C: 30,614	T: 14,350 C: 22,960
	0.90	T: 46,800 C: 84,240	T: 23,400 C: 42,120	T: 15,600 C: 28,080	T: 11,700 C: 21,060	T: 7,800 C: 14,040	T: 5,850 C: 10,530
	1.00	T: 26,800 C: 53,600	T: 13,400 C: 26,800	T: 8,934 C: 17,867	T: 6,700 C: 13,400	T: 4,467 C: 8,934	T: 3,350 C: 6,700

The required total number of PIT-tagged fish released at or above Lower Granite Dam to achieve the numbers required in Table A-1 will depend on how treatment and control groups are constructed. For example, in recent years around 10% to 15% of in-river fish have migrated undetected at Snake River dams and McNary Dam. Thus, if the control group for a particular test were to be made up of only never-detected fish, the total release for control group would be 8 to 10 times the “C” indicated in Table A-1. Alternatively, the never-detected group could be increased by modifying downstream dams or their operations; operating in primary bypass mode or having all guidance screens removed.

Duration and Intensity of Monitoring

Monitoring of PIT-tagged juveniles occurs automatically throughout the migration season at many sites. PIT-tagged adults are monitored automatically in the adult ladder at Lower Granite Dam. Results must be based on complete returns from each year’s outmigration, so monitoring of adults must continue three years after the last year of juvenile migration. Adult monitors will be installed at additional passage points above Bonneville Dam, and these may augment information currently obtained. It is necessary to install additional adult PIT tag detectors at strategic locations like Ice Harbor and Priest Rapids or Wanapum dams, to document any straying. Straying from LGR may depress TIRs and skew SARs to a particular basin, e.g., Snake versus upper Columbia.

A.1.3 Benefits, risks, costs, and trade-offs

Benefits and Amount of Learning Possible

Uncertainty exists about the benefit to different stocks of fish of transportation as presently implemented. Even less is known about return rates for fish with different juvenile passage histories. Continual monitoring will provide much needed information.

Considering projections of potentially greater adult return rates in the next few years, another 5 years of marking large numbers of juvenile fish will provide information for answering some broad-scale questions. For example, if the mean annual value of D is actually 0.8, another 5 years of data will very likely allow us to rule out the value 0.35. To distinguish between mean values of 0.7 and 0.8, however, would take much longer. Appendix F present some estimates of how long it would take to make such finer-scale distinctions with varying degrees of confidence.

Risks to Stocks

If transportation and/or the hydropower system have large impacts on fish, continual operation of the hydropower system and transportation will increase the risks that stocks will not recover. Direct risks to stocks would be minimal since recent studies have shown a benefit from transportation from Lower Granite Dam. Furthermore, by maximizing spill for inriver migrants, not all fish would be transported which would spread the risk between inriver migration and transportation as called for in the current Biological Opinion.

Costs

Transportation studies and the cost to mark stocks from all of the river basins to compare with upper river stocks will likely cost \$1.0 M + annually.

Trade-offs

There are few alternative to present hydropower system operations, other than dam removal. Recent sensitivity analyses showed that results of the decision analysis are quite sensitive to the value of D , while analyses of PIT-tag data suggested that recent and prospective D -values might be higher than previously estimated. Dam removal is a very costly alternative if it turns out that transportation does not currently impact stocks to the degree indicated by these previous estimates.

A.1.4 Inferences

If transported fish return at rates greater than the segment of fish that pass through the hydropower system under the best possible passage conditions and there are little or no delayed effects of transportation, it will indicate that transportation benefits fish. The degree that benefits are higher will affect the D value used in models.

A.1.5 Confounding factors

Measurements of transportation benefits are possibly confounded by numbers of hatchery fish in the system and conditions smolts face when entering the estuary and ocean.

A.1.6 Practical constraints

With regard to the general problem of estimating post-Bonneville survival for fish with different juvenile passage histories, there remain several important uncertainties regarding NMFS methods. The most

critical uncertainty is apparently the method of translating Lower Granite Dam-equivalent juvenile numbers into estimated numbers at Bonneville Dam, particularly regarding the proper survival probability to apply between McNary and Bonneville dams. Empirical survival estimates are not available for all species in all years. Where lower river estimates are not available, upper river estimates can be extrapolated, but the appropriate method of extrapolation is not clear. Where PIT-tag detection data below Bonneville Dam are available, the sampling methods (boat-towed trawl and recovery of tags deposited in bird nesting colonies) are very different from automatic monitoring in bypass systems. Detections below Bonneville Dam allow estimation of survival to the dam, but such detection data as is currently available must be used cautiously.

Methods for estimating the total number of tagged smolts arriving at Lower Granite Dam, which affects estimates of the never-detected group more than any other, are also under review.

With regard to the specific problem of estimating and testing D values, past uncertainty regarding proper weighting of SARs estimated for fish transported from different dams appears to be resolved. However, the issue may expose a more fundamental practical constraint. Though practical application may vary, D is defined in PATH models as a value that is applied to all transported fish in a given year, regardless of the dam at which they were collected. Moreover, D is modeled as a parameter that remains constant throughout the year, while recent PIT-tag data indicate that post-Bonneville Dam survival of transported fish can vary widely within a single season. Monitoring and testing can continue to be based on annual average estimates of D , but ultimately the most valuable insights may come from investigations of within-season patterns in post-Bonneville Dam survival. See Appendix F for further discussion of project-specific D values and between-year variation in D estimates.

A.2 Modify Transportation / Measure changes in SARs

B. Separate Wild/Hatchery Fish in Barges

Description of Experimental Action/Research & Monitoring

Rationale

Study Objective: Determine if reducing interaction of wild yearling chinook salmon with hatchery fish, particularly hatchery steelhead, during collection and transport from Snake River dams results in increased SARs.

Description of Hypothesis: Interaction of wild yearling chinook salmon with hatchery fish, particularly larger hatchery steelhead, during collection and transport from Snake River dams could be a contributing factor to the low SARs observed in recent years for transported fish. Increased hatchery production of smolts and increased collection at Snake River dams due to increased efficiency of collection systems has resulted in overcrowding of the facilities at times. The stress caused by this situation, as well as potential horizontal disease transmission, could lead to delayed mortality for both transported and non transported smolts resulting in reduced SARs.

Expected changes in SARs would range from no effect (treatment did not work) to elimination of all or a portion of the observed delayed mortality observed each year.

Experimental Action: This action requires continuing the current PIT-tag transportation experiments but with some portion of the release groups composed of wild spring chinook tagged and transported to below Bonneville Dam in isolation of steelhead. SARs for both control fish (i.e., fish transported under current operations) and treatment fish (i.e., fish tagged and transported with reduced steelhead

interactions) would be computed on a seasonal basis. Ideally, treatment and control groups could be transported in the same year, which would substantially increase the power of the test by eliminating year-to-year variability. Treatment and control groups could be released throughout the season (on separate barges) on a randomized basis or on the same barge in separate compartments.

Presumably, a minimum increase in SAR could be determined (say, 20%) that would justify continuing the program on a full time basis in the future. This increase in SAR could be tested for in a paired, one-sided test of the form:

$$H_0: SAR_T/SAR_C \leq 1.2 \quad [A5]$$

$$H_A: SAR_T/SAR_C > 1.2 \quad [A6]$$

where SAR_C is the SAR for the control group and SAR_T is the SAR for the treatment group.

Currently, the Lower Granite Dam juvenile collection facility does not have the ability to separate fish by species or size. Building a new juvenile facility at Lower Granite Dam, to include separation capabilities, has been discussed for many years, but has not been completed due to lack of agreement on design and pending decisions on transportation and dam removal.

The COE has funded the NMFS and the University of Idaho in recent years to evaluate potential separator designs including permanent primary separators and temporary secondary separators. Based on the results of these studies, either a permanent juvenile separator could be built at Lower Granite Dam or secondary separation methods employed within the existing bypass flumes, raceways, or both to separate wild yearling chinook salmon from larger hatchery steelhead. Temporary grading bars within the existing bypass flume leading to the transportation study raceways were successfully used to reduce handling of hatchery steelhead during marking for chinook salmon survival studies in past years.

Existing transport barges have separate compartments so that wild chinook salmon could be barged apart after separation and marking without greatly disrupting current transport operations.

Spatial and Temporal Components

Spatial contrasts are not possible with this approach. Temporal contrast could be generated by randomly releasing controls and treatment groups throughout the season.

Monitoring Approach

Variables to Monitor

The most important variable to monitor is SAR for control and treatment groups. For the optimal ability to contrast among treatment and control, a single SAR would be computed for each group per year. Temporal patterns could also be observed by computing SARs throughout the season.

In addition, it would be beneficial to examine samples of fish from both groups as they are released below Bonneville. Levels of stress could be determined by standard procedures. Also, behavioral differences in fish from the two groups could be detected by radio-tracking or other procedures.

Sample Sizes Required

Sample size requirements were determined using methods similar to those described in Section 3.1.1. The number of returning adults needed in each group (treatment and control) is calculated

$$n = \frac{2 \cdot (z_{(1-\beta)} + z_{\alpha})^2}{(\ln R^* - \ln R_0)} \quad [A7]$$

where R^* is the “true” ratio of SARs (SAR_T/SAR_C), R_0 is the ratio hypothesized under the null hypothesis, $(1 - \beta)$ is the power, and α is the significance level. Once the n is determined, SARs must be assumed to determine the number of smolts to tag for the treatment and control groups. Table A-2 provides yearly sample sizes for treatment and control groups under various assumptions. See Appendix F for discussion of additional factors affecting estimation of D over multiple years.

Duration and intensity of Monitoring

It would be important to release enough smolts per year such that a minimum level of detection could be achieved. It would be beneficial to conduct the experiment for a number of years to understand the year-to-year variability in the results and to strengthen any conclusions.

Benefits, Risks, Costs, and Trade-offs

Benefits and Amount of Learning Possible

Isolating the effects of interactions of hatchery stocks with wild fish during collection and transport will reduce the uncertainty about D and extra mortality for both transported and non transported fish. Also the action may result in new methods to increase the efficiency of transportation.

Risks to Stocks

The experimental actions outlined here would not increase risk to wild stocks since its is unlikely that co-mingling with hatchery steelhead provides any benefits.

Costs

Costs for separating wild and hatchery stocks at Lower Granite Dam would vary depending on the type of separator used. Options would range from installation of a permanent separator with design based on ongoing separator studies conducted at McNary and Ice Harbor Dams to installation of a temporary separator(s) for use within the existing system. Costs for PIT-tagging would be similar to transport studies described earlier.

Tradeoffs

Primary tradeoffs are that either more fish would have to be tagged in the transportation experiments or that some of the fish that would have been transported under the current conditions would be transported under the modified treatment conditions, potentially confusing results from T/I experiments. It might be possible, though, to conduct the regular T/I experiments, test for the steelhead interactions, and continue with methods to estimate the D parameter each year without modifying current experimental protocols substantially.

Another tradeoff is that this experiment would only detect interactions between steelhead and chinook during the collection and transportation operations. Detrimental interactions could occur in other phases of the life-history, including release to collection point and the estuary/ocean phase. While this experiment is tractable and could result in modifications of collection/transport operations, it will not fully characterize the full interaction between steelhead and chinook.

Table A-2: Numbers of PIT-tagged fish required yearly in treatment and control groups to detect hypothesized levels of effects of the treatment under various assumed SARs (for control fish), and hypothesized and true levels of the effect. The control group is fish transported under current operations. The treatment group is wild spring/summer chinook transported separate from steelhead. The ratio is of the SAR of the treatment groups to the SAR of the control groups. The significance level is $\alpha = 0.05$, and the power is $(1-\beta) = 0.80$.

Hypothesized Ratio = 1.2								
True Ratio	Adults needed		Expected LGR-LGR SAR for control					
			0.25	0.5	0.75	1	1.5	2
1.25	7421	cont.: treat.:	2968400 2374720	1484200 1187360	989467 791573	742100 593680	494733 395787	371050 296840
1.3	1930	cont.: treat.:	772000 593846	386000 296923	257333 197949	193000 148462	128667 98974	96500 74231
1.35	892	cont.: treat.:	356800 264296	178400 132148	118933 88099	89200 66074	59467 44049	44600 33037
1.4	521	cont.: treat.:	208400 148857	104200 74429	69467 49619	52100 37214	34733 24810	26050 18607
1.5	249	cont.: treat.:	99600 66400	49800 33200	33200 22133	24900 16600	16600 11067	12450 8300
1.6	150	cont.: treat.:	60000 37500	30000 18750	20000 12500	15000 9375	10000 6250	7500 4688
1.8	76	cont.: treat.:	30400 16889	15200 8444	10133 5630	7600 4222	5067 2815	3800 2111
2	48	cont.: treat.:	19200 9600	9600 4800	6400 3200	4800 2400	3200 1600	2400 1200

Hypothesized Ratio = 1.0								
True Ratio	Adults needed							
			0.25	0.5	0.75	1	1.5	2
1.1	1362	cont.: treat.:	544800 495273	272400 247636	181600 165091	136200 123818	90800 82545	68100 61909
1.2	372	cont.: treat.:	148800 124000	74400 62000	49600 41333	37200 31000	24800 20667	18600 15500
1.3	180	cont.: treat.:	72000 55385	36000 27692	24000 18462	18000 13846	12000 9231	9000 6923
1.5	76	cont.: treat.:	30400 20267	15200 10133	10133 6756	7600 5067	5067 3378	3800 2533

Inferences

One type of inference that would arise from these experiments is the ability to relate changes in SARs to Δm in the life cycle models. This would allow for an estimation of the impact of this management action in terms of projected stock levels.

Another consideration is how to conduct the experiment for several years. Two alternative approaches could be explored. In the first case, a “meta” analysis would be conducted where the single year hypothesis test is repeated over several years. This type of analysis would result in more confidence in conclusions compared to just a single trial. The second approach would be aimed more at characterizing year-to-year variability in the treatment/control ratios by estimating the mean and variance of the observed ratios over a several year period. In either case, further study is required to determine how many years the study should be repeated.

Confounding Factors

None unique to this experiment.

Practical Constraints

The methods of reducing interaction between wild and hatchery stocks during transport described here have no serious practical constraints beyond, perhaps, a limited ability to actually separate steelhead and chinook smolts.

A.3 Turn Transportation On/Off

NOTE:

Because we could not locate a precise description of the mechanism by which never-detected fish are thought to do better than fish detected many times, we have assumed that this is due to the secondary dewatering needed for detection (at least at transport projects) and subsequent transportation. While the methods described below would work for other mechanisms (e.g., bypass itself or primary dewatering) they are to some degree specific to the mechanism just noted.

A.3.1 Description of experimental action / research & monitoring

Background

Most unmarked (e.g., non PIT-tagged) Snake River chinook and steelhead smolts bypassed at collection projects are transported. In contrast, most marked fish bypassed at collector projects are returned to the river below the project outfall, to provide additional information for in-river survival studies. Therefore, for calculating SARs, TCR's, and other smolt-to-adult survival statistics, so-called “non-detected” PIT-tagged smolts – those never seen at collector projects – are often thought to be the best surrogates for the (unmarked) run-at-large for all Snake River ESU's.

Methodologically, this approach is not without potential problems (statistics are developed in Sandford and Smith 1999, in review). First, one must calculate the survival of non-detected fish from Lower Granite Dam (LGR) to each collection project. These calculations are very complex, in contrast to straight-forward Cormack-Jolly-Seber (CJS) survival estimates. In addition, since adult returns are sparse (due to low SAR's and small numbers released in each of many narrowly defined groups), the information content of the estimates of SAR's, TCR's, etc. is rather low. Because non-detected fish sometimes have higher SAR's than fish detected several times, there is the possibility that detection (i.e.,

bypass, dewatering, and detection in the lower reaches of each bypass system) may influence SARs. Under the hypotheses tested in this section, this long-term effect is assumed to result from secondary dewatering, rather than bypass proper. Finally, because relatively few wild smolts are tagged, and because few wild adults are scanned for PIT tags on the spawning grounds, calculating wild SARs to the spawning ground yields little useful information.

The mechanism by which conventional, PIT-tagged controls in T/C experiments may overstate the benefits of transportation is as follows. The SAR's for non-detected fish are sometimes somewhat higher for fish that are never detected in the bypass systems than for fish that are detected at several projects. This could be due to many factors, including imperfect detection systems. If the difference is real, it is attributed to the long-term effects of the PIT tag dewatering system. The mechanism is related to the treatment that the fish receives once it is in the bypass system. After a smolt is guided by the screens into the bypass system, a "primary" dewatering occurs. If no detection system is in place, the fish are then returned to the river at or below the project tailrace. Because the bypassed fish are in too large a pipe for accurate detection, fish at most projects with both detectors and transport facilities receive a secondary dewatering, and are shunted into a smaller pipe. There, they pass a series of magnetic coils designed to detect the tags. Fish that are to be transported are dewatered further before being placed in a raceway or directly into a barge. The secondary dewatering is thought to have deleterious long-term effects on the fish. Short-term effects (as measured by inriver survival through subsequent projects) are not apparent in studies done to date.

In this section, we develop rationale and an initial design for a large-scale adaptive management experiment to address these issues. If the design is both relevant (measures the intended effects) and sufficiently powerful, it will enable managers to obtain information on the smolt-to-adult survival of transported and non-transported fish without the complications noted above. We may also answer the question of whether or not transport, if fully applied to all Snake River stocks, could help those stocks to recover within a given time frame.

Rationale

- The objective is to determine if the SAR's, TCR's, and "D" values estimated in transport experiments are representative of the wild spring/summer and fall chinook runs at large. The method for doing so is to alternate different means for "treating" the population: in some years, most fish would be bypassed, dewatered, and transported, while in others nearly all fish would be bypassed but not dewatered or transported.
- The hypothesis to be tested is that measures of spawner-to-recruit survival are proportional to TCR's estimated from transport experiments.

A.3.2 Monitoring approach

The essence of this adaptive management experiment is to extend conventional, PIT-tag based experiments to include "true" controls, that would be more nearly representative of fish migrating in-river with little or no indirect influence of transportation. That is, in years with transport turned off (except for simultaneous transport experiments – see below), fish migrating inriver would be bypassed and primary-dewatered, but not secondary-dewatered. If secondary dewatering reduces subsequent survival, these "control" years should capture and measure the effect, whether it is evident from inriver survival downstream of transport projects, or in SAR's.

In addition, because the experiment would alternate years when most fish are transported with years when (almost) none would be loaded into barges, it should be possible to observe a strong contrast in measures

of survival that cannot be detected using PIT-tagged fish. In particular, we can monitor spawner-to-recruit survival across the two sharply different experimental conditions – transport vs. no transport – to see if these differences in survival comport with TCR's and other survival measures from conventional experiments.

Broadly speaking, there would be a “grand” adaptive management experiment – transport vs. no transport – with smaller, conventional transport experiments and inriver survival monitoring nested within the larger study (see Table A-3). The grand experiment would run in alternate years (See Hinrichsen appendix on why one should use even/odd years for this type of experiment). In years when transportation is in operation (call them T, T+2, T+4, ...) voluntary spill would be eliminated, all fish bypassed at transport projects would be dewatered and transported, while PIT tagging and detection would continue much like the present, though at higher intensity. In no-transport years (T+1, T+3, ...) projects would spill to gas caps, and detection systems would operate only on primary-dewatered fish (see above). Flat-plate detectors would be used to detect fish near the outfalls for these systems. Detection efficiency would likely be lower than at present (Bill Muir, pers. comm., Sept. 14, 1999; Jim Ceballos, pers. comm., Sept. 15, 1999); indeed, detection efficiency *could* be zero. Alternatively, one might route a portion (perhaps 10%) of the smolts through the secondary dewatering/detection apparatus. Collection/transport/tagging efforts at Lower Granite and other mainstem projects would operate only as needed to conduct transport experiments. Detection at John Day, The Dalles, and Bonneville would continue as at present in all years.

Table A-3: Operational and monitoring measures- transport-no transport adaptive management experiment

Operational Measures	Transport Years	No Transport Years
Spill	No Voluntary spill	Spill to gas caps
Screening/Bypass in place?	Yes	Yes
Dewatering of bypassed fish at transport projects	Yes	Minimal - just for marking and detection of experimental fish
Transport at Snake Projects (Granite, Goose, Lower Monumental)	Yes	Minimal - just for experimental fish
Transport at McNary?	Yes	Minimal - just for experimental fish
Bypass/detection below McNary?	Yes	Yes
Monitoring/Experimental Measures		
Collection and PIT tagging above Granite?	Yes, maximize	Yes, maximize
Collection and Marking at Granite	Yes	Only as needed for experimental transport SAR/TCR estimates
Collection and marking at other transport projects?	Yes	Only as needed for experimental transport SAR/TCR estimates

Monitoring would focus on many of the same variables as at present, including SAR's, TCR's, inriver survival (V_n), etc. (See Table A-4). One crucial addition would be that spawner-recruit survival could then be divided into years when fish are transported and years when most fish (transport study fish are the only exception) are transported.

Table A-4: Variables to monitor/estimate for transport-no transport experimental management action.

Variable	How to Monitor	
	Transport Years	No-transport Years
V_n	As at present	Use flat plates after primary dewatering. Detection probability likely lower
SAR - Project-Ocean-LGR	As at present	As at present. Smolt detection probability likely lower.
SAR - Project-Ocean-BONN	As at present (after BONN adult detectors in, 2001	As at present (after BONN adult detectors in, 2001
Spawner-Recruit Survival	As at present	As at present

To accurately assess the power of transport/no transport experiments, estimates of some variables may need to be improved. Note that the list is not exhaustive.

- 1) **Recruits:** For spring/summer chinook, these are recruits to the Columbia. For fall chinook, this will include ocean harvest. For retrospective work, we did not considered variance in the sub-model that is needed for these, except (for spring/summer chinook) potential errors in spawner estimates. We believe that we should include the recruitment variation, as best we are able, for prospective work. Details on harvest rate calculations, turnoff to tributaries, dam counts (including fallback rate estimates), and background data used to derive them will be required. If more PIT-tagged adults are detected at Bonneville in future, this information could substitute for the laboriously derived conversion rates used retrospectively.
- 2) **Spawners:** As with recruits, we believe we will need a more explicit treatment of variation in the estimates for prospective work. As part of this, we will need details on how dam counts (for fall chinook) and redd counts (for spring/summer chinook) are used to calculate spawning abundance. This includes redd-count expansion factors and their derivation, hatchery contributions to natural spawning, aging data and expansions to run-year spawner estimates, pre-spawning mortality, their respective data sources, and assessments of the accuracy of the expansions/extrapolations.
- 3) **Passage parameters (P , V_n , and D):** We obviously assume that these will be calculated from PIT tag release/detection information. Analytical or boot-strapped variance estimates can and should be calculated for all these quantities. Explicit assumptions relating PIT tag population estimates to individual stocks will also be required. Finally, for the Delta-style model analog of Equation [A8], estimates of V_n and perhaps SARs will be needed for downstream spring chinook stocks.

Example Analysis: Precision of In-river Survival Estimates [V_n]

V_n (in-river survival of non-transported fish) and $P(bt)$ (proportion of fish arriving below Bonneville Dam that were transported) estimates are needed to assess changes in extra mortality, in addition to being of interest in their own right. $P(bt)$ is obviously not a problem in years when only experimental fish are transported: it will probably be close to zero. V_n estimates depend on detections of tagged smolts at dams,

and these detections would decrease dramatically in years with reduced secondary dewatering. We assume that the number of fish tagged is roughly equal in years with transport operating and years when only experimental fish are transported. Furthermore, we assume that at each transport project, the same number of fish will be tagged and released for transport experiments regardless of what occurs for the run at large. This is purely for convenience in subsequent power analyses. It may turn out to be a sub-optimal strategy, in the sense that one could gain more information by varying the tagging protocol between transport/no transport years.

In years T, T+2, etc. when transportation is operating, estimates of inriver survival would proceed in much the same way as at present. Most PIT-tagged fish would not be transported, but used to estimate inriver survival. They would therefore be returned to the river at each project. The exception would be fish selected for transport experiments. By assumption, the transport experiments would occur at each project where fish are transported, with roughly half of the fish being tagged (if they aren't already) and placed into barges, and remaining half returned to the river at outfalls below the project.

In years T+1, T+3, etc. estimates of inriver survival will be less precise, since detection efficiency will surely decrease. This will occur because detecting fish in larger primary bypass outlets will be less efficient than in secondary-bypass outlets, if it is even possible. Finally, since spill is increased in no-transport years, a higher proportion of the smolts will be directed away from the bypass and detection facilities. Note that the tagging protocol at transport projects would also need to be changed: since a smaller portion of the run will go into secondary bypass/detection/collection systems, obtaining a fixed number of fish for a transport experiment may take longer than under current operations.

To see how this might affect the variance in in-river survival estimates, we conducted a very simple “experiment” using chinook (from PITAGIS) that were released above (not at) Lower Granite in 1998, with all fish having the tagger-assigned migration year of 1998. The experiment was to randomly reduce detections at Lower Granite, Little Goose, Lower Monumental, and McNary (i.e., all transport projects) by 90%, and see how this affected estimates of the standard errors for in-river survival estimates.

In doing so, we used a very simple approach: spring, summer, and fall chinook of hatchery origin comprised three release groups, while wild fish comprised three additional groups. That is, we did not try to separate fish by day, week, or other sub-seasonal release or detection time. Note that we do not recommend this as a strategy to estimate seasonal average survivals; rather, it is a way of developing a simple example. If, with reduced detections, one could form the approximately the same number of daily or weekly groups as at present, the method we used will overstate the increase in variance associated with a reduction in detections.

Data used for the example are shown in Table A-5. Releases included only fish of known run and origin, from release sites well above LGR. As can be seen, the numbers of releases vary by almost three orders of magnitude, from 162K for hatchery spring chinook to 556 for wild fall chinook. Detections follow a similar pattern. We did not include “detections” at Rice Island, as these tend to be very modest numbers for chinook.

Table A-6 shows CJS estimated survival and detection rates, both for “All” detections and the 90% reduction, as described above. As one would expect, point estimates of survival generally don't change by much – where they do, we believe it's simply due to the randomness in detection rates. We only performed each simulated reduction in detections once; had we repeated it more often, survival rates would have converged to the means for the “All” case. A 90% reduction on detection rates usually increases the standard error (S.E.) of the survival rates by a factor of 10. For example, for hatchery fall chinook, survival from release to LGR was about 0.64. When all detections are used, the S.E. is 0.003. When 90% of detections are ignored, the S.E. increases to 0.027.

Table A-5: Releases and Detections, Chinook Released and Migrating in 1998

Chinook Run	Hatchery/Wild	Event	Number Released or Detected
Spring	Hatchery	Released Above LGR	162,021
		Lower Granite Detections	52,532
		Little Goose Detections	35,440
		Lower Mon. Detections	23,756
		McNary Detections	15,165
		John Day Detections	6,563
		Bonn. Detections	4,999
		Towed Array (Below Bonn.) Detections	1,125
Summer	Hatchery	Released Above LGR	49,404
		Lower Granite Detections	13,473
		Little Goose Detections	9,506
		Lower Mon. Detections	7,441
		McNary Detections	3,168
		John Day Detections	2,079
		Bonn. Detections	1,473
		Towed Array Detections	228
Fall	Hatchery	Released Above LGR	102,596
		Lower Granite Detections	30,002
		Little Goose Detections	26,394
		Lower Mon. Detections	17,028
		McNary Detections	13,902
		John Day Detections	3,988
		Bonn. Detections	1,302
		Towed Array Detections	89
Spring	Wild	Released Above LGR	5,511
		Lower Granite Detections	1,451
		Little Goose Detections	1,450
		Lower Mon. Detections	1,090
		McNary Detections	606
		John Day Detections	391
		Bonn. Detections	206
		Towed Array Detections	22
Summer	Wild	Released Above LGR	6,110
		Lower Granite Detections	2,316
		Little Goose Detections	2,417
		Lower Mon. Detections	1,595
		McNary Detections	1,063
		John Day Detections	470
		Bonn. Detections	300
		Towed Array Detections	47
Fall	Wild	Released Above LGR	556
		Lower Granite Detections	124
		Little Goose Detections	121
		Lower Mon. Detections	68
		McNary Detections	58
		John Day Detections	27
		Bonn. Detections	11
		81	

Table A-6: “Seasonal” CJS Survival and Detection Estimates, Chinook Released and Migrating in 1998. Reported survival rates are for the reach indicated in the column header; reported capture rates are at the last dam in that reach. For example, in the first column of numbers, survival rates are from release to LGR, while capture rates are at LGR.

			Towed Array Detections				0						
Chinook Run	Hatchery/Wild	All or 10% Detections		Release to LGR (Capture at LGR)	(S.E.)	LGR to LGS (Capture at LGS)	(S.E.)	LGS to LMN (Capture at LMN)	(S.E.)	LMN to MCN (Capture at MCN)	(S.E.)	MCN to JDA (Capture at JDA)	(S.E.)
Spring	Hatchery	All	Est. Survival	0.728	(0.002)	0.984	(0.006)	0.852	(0.007)	0.886	(0.013)	0.784	(0.025)
			Est. P (Capture)	0.445	(0.002)	0.44	(0.003)	0.386	(0.003)	0.278	(0.004)	0.154	(0.005)
Summer	Hatchery	All	Survival	0.595	(0.004)	0.957	(0.010)	0.861	(0.012)	0.885	(0.027)	0.822	(0.048)
			Capture	0.459	(0.004)	0.482	(0.005)	0.466	(0.007)	0.224	(0.007)	0.178	(0.010)
Fall	Hatchery	All	Survival	0.636	(0.003)	0.777	(0.004)	0.901	(0.008)	0.822	(0.015)	0.77	(0.051)
			Capture	0.46	(0.002)	0.529	(0.003)	0.385	(0.004)	0.382	(0.007)	0.142	(0.009)
Spring	Wild	All	Survival	0.586	(0.009)	0.946	(0.019)	0.942	(0.032)	0.94	(0.068)	0.7	(0.095)
			Capture	0.45	(0.010)	0.494	(0.012)	0.409	(0.015)	0.242	(0.018)	0.223	(0.028)
Summer	Wild	All	Survival	0.781	(0.008)	0.991	(0.015)	0.874	(0.024)	0.89	(0.048)	0.803	(0.098)
			Capture	0.485	(0.008)	0.531	(0.010)	0.414	(0.012)	0.309	(0.017)	0.171	(0.020)
Fall	Wild	All	Survival	0.566	(0.043)	0.639	(0.059)	1.089	(0.160)	0.484	(0.098)	0.48	(0.141)
			Capture	0.395	(0.038)	0.604	(0.045)	0.319	(0.055)	0.563	(0.088)	0.545	(0.150)
Spring	Hatchery	10%	Survival	0.524	(0.023)	0.947	(0.057)	0.796	(0.044)	0.963	(0.060)	0.789	(0.045)
			Capture	0.062	(0.003)	0.046	(0.002)	0.038	(0.002)	0.027	(0.002)	0.142	(0.005)
Summer	Hatchery	10%	Survival	0.415	(0.032)	0.936	(0.094)	0.773	(0.065)	1.026	(0.121)	0.855	(0.098)
			Capture	0.063	(0.005)	0.053	(0.004)	0.057	(0.004)	0.02	(0.002)	0.167	(0.009)
Fall	Hatchery	10%	Survival	0.634	(0.027)	0.788	(0.049)	0.916	(0.069)	0.703	(0.060)	0.859	(0.076)
			Capture	0.046	(0.002)	0.051	(0.003)	0.037	(0.002)	0.042	(0.003)	0.141	(0.009)
Spring	Wild	10%	Survival	0.499	(0.063)	1.121	(0.227)	0.985	(0.249)	0.734	(0.206)	0.799	(0.185)
			Capture	0.059	(0.009)	0.046	(0.008)	0.032	(0.007)	0.032	(0.007)	0.222	(0.028)
Summer	Wild	10%	Survival	0.84	(0.10)	0.81	(0.14)	0.87	(0.17)	1.05	(0.26)	0.73	(0.16)
			Capture	0.046	(0.006)	0.057	(0.008)	0.042	(0.007)	0.027	(0.006)	0.172	(0.020)
Fall	Wild	10%	Survival	0.265	(0.105)	1.088	(0.674)	1.488	(1.532)	0.269	(0.279)	0.773	(0.422)
			Capture	0.082	(0.039)	0.075	(0.042)	0.029	(0.029)	0.063	(0.043)	0.545	(0.150)
Spring/Summer	Wild	10%	Survival	0.776	(0.079)	0.775	(0.107)	0.847	(0.118)	1.098	(0.200)	0.702	(0.120)
			Capture	0.041	(0.005)	0.052	(0.006)	0.048	(0.006)	0.026	(0.004)	0.19	(0.016)

While the results for hatchery fish suggest that the precision of survival estimates would still be fairly tight for reaches with good estimates at present, for wild fish the results are more discouraging. For example, for wild spring chinook survival from LGR to LGS, estimated survival using all detections is 0.946, with a S.E. of 0.019. If the detection rate is reduced by 90%, the point estimate changes to 1.121, with a S.E. of 0.227. However, the example likely overstates the problem, at least for spring/summer chinook. Aside from the over-simplified nature of this “seasonal” estimate already noted, spring and summer chinook are usually combined (see last section of Table A-6). In addition, many wild fish are tagged in their subbasin of origin the previous year; a back-of-the envelope estimate from Section 3.5 suggests that about 20K fish were tagged in 1997, and presumably migrated in 1998. If 25% of these fish survived to LGR, it would add at least 5K fish to the available sample. Finally, PIT-tagged wild fish may well be needed for other studies. Assuming that sufficient wild parr can be found, it seems reasonable to assume that the sample sizes used in this example are most likely under-estimates.

In summary, it would appear from this example that detection rates could be reduced substantially without greatly reducing the precision of in-river survival estimates.

Problems Using Historical R/S Models for Power Analyses

If one wishes to be able to estimate the power of tests for changes in R/S between transport and no-transport years, it seems logical to use a model of historic R/S to see how much unexplained noise is in

that data. The retrospective model could then be used to assess the power of various experimental designs. For this example, we used the Alpha-style model, for three reasons:

- 1) It is simpler than the Delta model, with no year effects;
- 2) It makes no assumptions about common ocean survival for the upstream and downstream stocks; and
- 3) It can be translated more easily into a model that could be used for fall chinook.

We used a simple (for PATH) version for spring/summer chinook:

$$\text{Ln}(R) = \text{Ln}(S) - M - \ln(D \cdot P + 1 - P) - b_0 * S + b_1 * \text{Step} + b_2 * 1/F + b_3 * P/F + \text{epsilon} \quad [\text{A8}]$$

Where

$\text{Ln}(R)$ = Natural log of recruits;
 $\text{Ln}(S)$ = Natural log of spawners;
 M = Total Passage mortality (from summer 1997 versions of CRISP and FLUSH);
 D = post-Bonn. survival of transported fish relative to in-river fish;
 P = Proportion of fish below Bonneville that were transported there;
 S = Spawners;
 Step = 0 through brood year 1974, 1 thereafter;
 F = Flow at Astoria, in year of downstream migration;
 P = Poppa Drift, 1st winter of ocean life;
 epsilon is a normally distributed mixed process and measurement error; and
 $b_0 - b_3$ are estimated parameters.

Note that in this formulation, $\text{Ln}(S)$, $-M$, and $\ln(D \cdot P + 1 - P)$ are “offsets,” with parameters equal to minus one, by definition. As such, they do not “count” in terms of the number of estimated parameters in the model.

As formulated above, the passage models’ estimates are assumed to account for both direct (i.e., in-river survival) and indirect (delayed transport) effects of passage through the hydrosystem. We were concerned because retrospective estimates of these terms are derived from passage models. Since prospective estimates will be derived from PIT tag data, we might, in effect, be accounting for the same phenomena (in-river survival, $P[\text{bt}]$, and “ D ”) using two different methods. This in turn could affect the power analysis in ways that are difficult or impossible to account for.

However, in the process of developing the retrospective R/S analysis, we have simply assumed that passage terms are important in explaining recruitment. PATH has never systematically tested this assumption. More explicitly, we have never compared the performance of models like Equation [A8] to a slightly simpler model:

$$\text{Ln}(R) = \text{Ln}(S) - b_0 * S + b_1 * \text{Step} + b_2 * 1/F + b_3 * P/F + \text{epsilon} \quad [\text{A9}]$$

With all terms as defined above. Before worrying about how to reconcile retrospective and prospective estimates of passage survival, we decided to compare the two. The results are shown in Table A-7.

Table A-7: Comparison of Goodness-of-Fit (R-Square) for Alpha-style Models With and Without Passage Model Offsets. CRISP and FLUSH Results Are From August, 1997.

	Deviance (SSE)	R-Square
Null Model [$\ln(R) = \text{mean } \ln(R)$]	524.280	0.000
Eqn. A.3.2 (No Passage Offset)	219.193	0.582
Model A.3.1 (With Passage Offset):		
CRISP C1	200.900	0.617
CRISP C3	205.285	0.608
CRISP C4	203.423	0.612
FLUSH F1	203.924	0.611
FLUSH F3	211.210	0.597
FLUSH F4	207.585	0.604

As can be seen from the table, adding the passage offsets (Equation [A8]) results in only slightly better Sum-of-Squares (SSE's) and R-squares than the simpler model with no offsets (Equation [A9]).¹¹ Given the $\approx 40\%$ of variation that is unexplained by the models (i.e., $1 - R\text{-Square}$), and the difficulties reconciling retrospective and prospective passage parameter estimates already noted, we conclude that, for experimental management planning, we will be better off simply ignoring the passage models. Furthermore, we conclude that the "historic," pre-PIT-tag era stock-recruit data will not add much information to the transport/no transport experimental management alternative. Instead, we should concentrate on obtaining precise, accurate estimates of data and parameters in future.

SAR Variability Example and a Cautionary Note

As an example of the variability of SARs and TCRs, we examined data for spring/summer chinook tagged at Lower Granite (LGR) in 1995-1996. Data in the top half of Table A-8 are for fish tagged as part of the NMFS transport study in each year, with the last column being 1995 and 1996 combined. The first line displays the number transported, and the second the number returned to the river below the project (fish transported at other projects are excluded from the sample). The next two lines show jack and adult returns to LGR through October 7, 1999. [The 1999 returns should account for most 3-ocean fish that migrated downstream in 1996.] The SAR's in the next lines are simply (adults + jacks) / releases for each group, and their standard errors (Burnham et al 1987, p. 115). The TCR is the ratio of the two SAR's. It's variance is computed using Burnham et al p. 84. As can be seen from the table, the standard errors are fairly tight on all survivals and TCRs. More importantly, SARs for 1995 are 3-4 times higher in 1995 than in 1996.

¹¹ Note that this is analogous to results for fall chinook, where adding passage model parameters did not make much difference in goodness-of-fit measures.

Table A-8: SAR's and TCR's for 1995 and 1996 Spring/summer chinook transport studies at Lower Granite (LGR). Data from PITAGIS; Doug Marsh study groups tagged at LGR. Includes fish detected below LGR. Excludes fish transported below LGR.

	Migration Year		
	95	96	95 & 96 Combined
# Transported @ LGR	101,576	44,799	146,375
# In-river @ LGR	125,070	64,578	189,648
Transported Jacks + Adults @ LGR	516	57	573
In-river Jacks + Adults @ LGR	331	54	385
Transport SAR	0.508%	0.127%	0.391%
Std. Dev.	0.022%	0.017%	0.016%
In-river SAR	0.265%	0.084%	0.203%
Std. Dev.	0.015%	0.011%	0.010%
TCR	1.92	1.52	1.93
Std. Dev.	0.135	0.289	0.127
Between-Year Comparisons			
	95 in-river data, 96 transport data	96 in-river data, 95 transport data	95 & 96 Combined
# Transported @ LGR	44,799	101,576	146,375
# In-river @ LGR	125,070	64,578	189,648
Transported Jacks + Adults @ LGR	57	516	573
In-river Jacks + Adults @ LGR	331	54	385
Transport SAR	0.127%	0.508%	0.391%
Std. Dev.	0.017%	0.022%	0.016%
In-river SAR	0.265%	0.084%	0.203%
Std. Dev.	0.015%	0.011%	0.010%
TCR	0.48	6.08	1.93
Std. Dev.	0.069	0.868	0.127

Recall that in the transport/no transport experiment, one would like to compare SARs and R/S in years with and without transport. Contrast this with a "within-year," conventional transport experiment. In a conventional experiment, the SARs derived for each group may be thought of as follows.

$$\text{SAR}(\text{in-river}) = V_n * (\text{Bonneville to LGR survival}) \quad [\text{A10}]$$

$$\text{SAR}(\text{Transport}) = 0.98 * D * (\text{Bonneville to LGR survival of in-river fish}) \quad [\text{A11}]$$

Where V_n = in-river survival;
 0.98 = assumed in-barge survival of transported fish;
 D = Survival differential for transported fish, compared to in-river migrants;
 Bonneville to LGR survival = “Common” survival for both groups, Bonneville back to LGR, after allowing for “D”. This is similar to method used for PATH retrospective “D” calculations.

It is straight-forward to solve for the common survival for each year. Assuming in-river survival of 0.5, the common survival is about 0.5% for 1995 and 0.25% for 1996. Note that the common survival [$\text{SAR}(\text{in-river}) / V_n$] is determined solely by in-river migrants. Because the TCRs and “D” values are ratios containing the common term, [e.g., $\text{SAR}(\text{transport})/\text{SAR}(\text{in-river})$], the common survival cancels out for within-year comparisons.

For the transport/no transport experiment, however, one would compare recruits/spawner between years with and without transportation: a between-year comparison. To see what the effect of this might be, in the lower half of Table A-8, we “crossed” the data for 1995 and 1996. In the first column of the second half of the table, we used data for 1995 in-river migrants (as a stand-in for a no-transport year) and 1996 transported fish. In the second column of the second half, we reversed the comparison, using 1996 in-river migrants and 1995 transported fish. The results, as one might expect given very different “common” SARs for the two years (see above) were very discouraging: the pseudo-TCRs were 0.48 and 6.08 for the two cross-year comparisons.

An Example Power Analysis Using 1980-90 R/S Estimates and Assumed T/C Relationships

Given this high variability in Bonneville to LGR survival and the large amount of unexplained noise in the R/S data series noted previously, one may need to carry out between-year comparisons for many years before being able to detect differences in R/S between transport and in-river years. However, if the two-fold difference in SAR’s in transport experiments (e.g., 1995 data from Table A-8) carries over to R/S, some more rapid results may be possible. Unfortunately, the R/S data presently available (through brood year 1990 for spring/summer chinook, and brood year 1991 for fall chinook) ends at about the same time that PIT-tag based SAR estimates begin, around 1993-1995. Therefore, we cannot establish whether there is any correspondence between R/S and SAR’s based on PIT tags. However, there was a weak correlation between Raymond’s SAR estimates and R/S for spring chinook (e.g., PATH FY98 report), so there is some reason to believe that such a relationship should hold at present. One way to test this assumption would be to update the R/S analysis for spring chinook (through brood year 1994, downstream passage year 1996) and perhaps fall chinook (through brood year 1993, passage year 1994), and compare R/S to available PIT-tag SAR estimates. However, because the overlap between R/S and SAR data are very short, one cannot expect too much from such a comparison.

By way of an example with existing data, we have developed a very simple stock-recruit model. It uses spring/summer chinook R/S data from 1980-90, after transport began in earnest. The retrospective model has the following form:

$$\text{Ln}(R_{j,t} / S_{j,t}) = \text{YEAR}_t + \text{Epsilon}_{j,t} \quad [\text{A12}]$$

Where:

$R_{j,t}$ = recruits, stock j , year t
 $S_{j,t}$ = spawners, stock j , year t
 $YEAR_t$ = Year effect factor or class variable, year t , and
 $Epsilon_{j,t}$ is mixed process and measurement error.

No density-dependent term is included, since spawner densities in 1980's were very low.¹²

Now, what one would like to do is see how R/S varies with the proportion of the run transported. Unfortunately, according to passage model output, the P_{bt} was very high throughout this period: since there is little contrast in the existing data, we cannot use it to test hypotheses about TCR's, etc. Instead, for purposes of an example analysis, we will assume that differences in TCR's do translate into differences in R/S , and perform a simple bootstrap analysis to see how long it would take to detect the (assumed) differences.

More precisely, assume that in future the following relationships hold:

In transport years:

$$\ln(R_{j,t} / S_{j,t}) = YEAR_t + TRANS_T + Epsilon_{j,t} \quad [A13]$$

While in non-transport years

$$\ln(R_{j,t} / S_{j,t} / TCR) = YEAR_t + TRANS_T + Epsilon_{j,t} \quad [A14]$$

Where $TRANS_T$ is a dummy variable, 1 when transport is operating and zero otherwise. In the absence of empirical data, Equation [A14] essentially assumes that the TCR's carry through to recruitment: the higher the TCR, the lower R/S should be in years when transport is turned off. The question then becomes: how long would one need to detect such differences, assuming the above relationship is in fact correct?

Table A-9 contains the results from a simple 50-iteration bootstrap test. We have used the conventional 0.8, 0.05 combination often employed in biological power testing: an experiment is sufficiently powerful if it detects the effect of interest 80% of the time with 5% confidence limits. Shaded cells indicate sufficient power as just defined. As one can see, TCR's of 1.2 will not be detectable within 11 years. Note that this is 11 years of different passage treatments, plus 3-4 years for all recruits to return, for a total of 14-15 calendar years from the start of the experiment. Conversely, if the TCR is 2.0, this would be detectable within 3 years 95% of the time.

¹²

In fact, when a model is estimated that includes spawners, the coefficients are usually significant and positive: more spawners is associated with higher R/S . This lends some support to the notion that spawner densities may be "too low" at present, and that adding nutrients could enhance survival.

Table A-9: Example power analysis for a range of TCR and monitoring periods. Shaded cells indicate power of at least 80% at 5% Type-1 error level. See text for details.

Probability of Detecting an Effect @ 5% Type-1 Error Level						
Assumed TCR	Years of Monitoring					
	1	3	5	7	9	11
1	0	0	0	0	0	0
1.2	0	0	0	0	0	0
1.4	0	0	0.16	0.26	0.46	0.8
1.6	0	0.33	0.48	0.67	0.86	0.98
1.8	0.21	0.53	0.75	0.92	0.98	1
2	0.71	0.95	0.98	1	1	1

There are a number of reasons why the above example will overstate the power of the experiment. Firstly, it assumes the relationship that it purports to test: that TCR's are directly related to spawner \rightarrow recruit survival. Given the poor fit between passage model survival estimates and R/S (above) this likely overstates the relationship. Secondly, due to time constraints, we have not accounted for some of the variability in R/S and the inter-annual variance in the TCR's themselves. Other problems will doubtless arise following more extensive review of the methods.

On the other hand, the example results may understate the power of the experiment. R/S estimates for 1980-90 are available for at least 9 additional index stocks (data from Ray Beamesderfer, August 20, 1997). Use of this information in models like Equations [A12] - [A14] would probably increase the power of future experiments. In addition, if we can perform T/C experiments concurrently with a transport/no transport experiment, that information may add to the confidence one can place in the experimental management results. This could be done by subsuming SAR estimates for transport and no-transport years (from PIT tag transport and control groups) into models similar to Equations [A12] - [A14].

A.3.3 Benefits, risks, costs, and trade-offs

Benefits and Amount of Learning Possible

The proposed experiment should greatly reduce the current uncertainties associated with the benefits (if any) of transportation. Unless an analysis prior to the experiment shows this to be the case, it will not be implemented.

Risks to Stocks

The obvious risk is that if transportation is beneficial, eliminating it for the run-at-large $\frac{1}{2}$ of the time will be an obvious problem. On the other hand, if we had complete certainty about the effects of transportation, we would not carry out the experiment in the first place.

Costs

Costs should be roughly the same as for current transport experiments. Additional tagging efforts would increase costs, but this will be offset (to some degree) by reductions in spill (i.e., foregone power costs) and transportation in years when these are reduced.

*Trade-offs***A.3.4 Inferences**

- Short-term operations (3-10 years): If the adaptive management experiment shows that past transport experiments' results (TCR's approximately equal to 2) provide an accurate depiction of the relative benefits of transportation, then current "spread the risk" operations would be replaced by operations that maximize transportation. In contrast, if the adaptive management experiment suggests that "true" TCR's are less than one, one would discontinue transportation.
- Long-term operations (11+ years): if the SAR's from the adaptive management experiment suggest that, in combination with other measures, survival with full transportation is sufficient to lead to recovery, maximize transportation. Otherwise, proceed with 2-dam or 4-dam drawdown.

A.3.5 Confounding factors

Obviously, any other experiments that were carried out concurrently (e.g., fertilization, reduced hatchery releases) might confound the results. The key to reducing the confounding is to vary the other experiments on a different schedule than that used for the transport experiment.

A.3.6 Practical constraints

None that are unique to this experiment. So far as we know, it could be carried out under existing regulatory and legislative frameworks.

A.4 Carcass Introductions / Stream Fertilization**A.4.1 Description of experimental action / research & monitoring**

Description: Life cycle survival is reduced because there are too few spawner carcasses to provide adequate nutrients in natal and freshwater rearing areas. This may be manifest as a decrease in parr-smolt survival or spawner(t)-spawner(t+1) survival. Either may be due in part to reduced parr or smolt size.

Experimental Action: Experimental carcass introduction or introduction of chemical fertilizers to increase stream nutrient levels. As nutrients increase, then parr-smolt mortality, and perhaps "extra" spawner-recruit mortality will decrease. Parr in about 30 rearing areas are already PIT-tagged, about 16 of which have data for six of the past seven years. Therefore, there are opportunities for staircase-style experimental designs for both parr-smolt and R/S monitoring (see below).

Evidence against this hypothesis:

- No change in smolts/spawner since 1960's (Chapter 9 of FY96 PATH report).
- No evidence of depensation in spawner-recruit data (estimated depensation parameter, "p" = 0).
- Lemhi (higher nutrients) has shown a rate of decline similar to other stocks.

Counter-arguments:

- The evidence in Chapter 9 depends on two different methods for estimating smolt abundance between the early 1960's and the 1980's. This may reduce the confidence one can place in the inferences.

- Even if there is no significant difference in smolts/spawner from 1960's, there may be physiological effects that cause smolts to be less "fit" since spawner numbers decreased, which affects recruitment. This may be manifest as a reduction in smolt length and/or weight.
- The apparent lack of depensation in stock-recruitment data could occur because of insufficient data points at low spawner abundance.
- Measurement error may conceal depensation (Hinrichsen, *in prep*).

Other evidence:

- Kline et al (1990) show that marine-derived nitrogen and carbon are recycled by stream biota (this is one of numerous examples of similar work in Alaska, British Columbia, and Washington).
- Johnston et al (1990) and Stockner et al (1996) exemplify work on lake enhancement in British Columbia.
- Bilby et al (1996) demonstrated that marine-derived nitrogen and carbon from coho carcasses are incorporated into stream biota, including coho smolts. They also showed that growth rates of age-0 coho doubled following spawning (in their 2nd year of freshwater rearing).
- Bilby et al (1998) concluded that age-0 coho and age-0/1+ steelhead densities increased following the addition of coho carcasses.
- Michael (1995) demonstrated a strong positive correlation between the abundance of pink salmon spawners and recruitment of coho rearing in the same streams in the year the pinks spawned.
- There is an extensive literature on the incorporation of marine-derived nutrients into stream biota, including age-0/1+ anadromous sockeye, steelhead, and coho, and resident trout. Evidence of increases in density and size of parr/smolts also exists, but is not so extensive.
- Michael (1995) is apparently the only study that carries survival through to adult recruitment (but see Schmidt et al 1998 for a more indirect approach using sockeye).
- No similar studies have been done on chinook, although one by Bilby (1999) is starting this year. Because all studies to date are for salmonids other than chinook, the effects (if any) of carcass or nutrient additions are essentially unknown. The one exception is a recent (1998-99?) study in the Grande Ronde (N. E. Oregon, Howard Schaller, pers. comm., 7/26/99).

Spatial and Temporal Components

In about 32 sites in the Snake tributaries, rearing spring/summer chinook parr are already PIT-tagged in the summer and fall (Table A-10). Survivors are detected the following spring at traps and mainstem dams on the Snake and lower Columbia. Many of the sites (e.g., Bear Valley and Elk Creek) are probably too close geographically to use as separate experimental sites (enhancement in one creek would likely have similar effects on both), but a substantial number of well-separated sites should be feasible. At 16 sites, fish have been tagged in 6-7 of the past 7 years (see Table A-11). Mean survival (naïve bootstrap, 5,000 draws from individual tagging/detection records) varies widely among sites and years (Table A-13). Length of tagging is almost always recorded (Table A-12), and the variance in the Cormack-Jolly-Seber estimates of overwintering survival is generally modest. This is true especially in later years (1994-on, with several monitoring sites at mainstem dams and most tagged fish returned to river), and where > 1,000 parr were tagged (Table A-13, Figure A-1). In general, there is a marked decrease in the "range" of survival estimates (defined here as [95th percentile- 5th percentile] / median survival) as the number of fish tagged approached 1000-2000, with much smaller decreases thereafter. From this, we conclude that increasing sampling effort to obtain 2000 +/- fish at each site and year would increase the precision of survival estimates (from tagging as parr in the summer/fall to LGR the following spring), but that samples > 2k would add little additional information. Power analyses are performed (Section A.4.2.2, below) assuming no increase in tagging effort, however.

Table A-10: 32 Sites with tagging data and Number of fish tagged, 1992-98.

Site	1992 # Tagged	1993 # Tagged	1994 # Tagged	1995 # Tagged	1996 # Tagged	1997 # Tagged	1998 # Tagged
Altulc	368	-	331	-	-	-	-
Bear/Elk	1632	1854	2916	-	-	671	1519
Bigc	758	730	1499	-	-	-	1452
Camasc	1011	215	1527	4	-	-	-
Capehc	205	-	1326	-	-	-	270
Cathec	1091	998	1983	1102	982	1250	1151
Cfctrp	855	1857	2883	359	538	988	2618
Chambc	497	570	1157	-	-	-	-
Crotrp	84	357	1164	40	-	84	273
Fren/Smile	541	892	1103	500	-	-	-
Grandr	915	1909	1853	-	27	724	937
Imnahr	996	2427	1758	2973	1458	4421	5003
Johnsc	633	-	192	-	-	-	5444
Lemhir	560	746	1717	179	269	752	3463
Loloc	923	1503	1639	144	-	620	2003
Lookgc	-	1944	3569	2025	15	1626	2151
Loonc	261	395	964	-	-	-	1030
Lostir	995	721	999	977	1045	997	1172
Marshc	999	7534	4891	275	-	1006	2971
Minamr	935	994	996	988	589	984	999
Pahsir	1072	561	2928	262	101	248	1160
Red/Amer	552	996	2758	634	25	1385	1571
Salref	222	316	1576	108	-	-	960
Salrnf	505	314	519	-	-	-	-
Salrsf	640	5196	3999	1777	2048	2869	3920
Sawtrp	739	99	1132	553	-	116	351
Secesr	-	673	1547	571	260	1176	3033
Sulfuc	710	-	726	-	-	-	442
Valeyc	1026	848	1550	-	-	-	1001
Wenr/Wenrsf	730	995	996	993	62	-	-

Table A-11: Site names, locations, and climate regions for 16 sites with 6 - 7 years of tagging data, 1992-1998.

PITAGIS Site ID	Name	Palmer Drought Severity Index (PDSI) Climate Region
Cathec	Catherine CK - OR	NE_OR
Cfctrp	Crooked Fork Trap - ID	N_Cent_Canyons
Crotrp	Crooked Trap - ID	N_Cent_Canyons
Grandr	Grande Ronde - OR	NE_OR
Imnahr	Imnaha - OR	N_Cent_Canyons
Lemhir	Lemhi - ID	NE_Valleys
Loloc	Lolo Ck - ID	N_Cent_Canyons
Lookgc	Looking Glass CK - OR	Blues
Lostir	Lostine - OR	NE_OR
Marshc	Marsh Ck - ID	Cent_Mts
Minamr	Minam - OR	NE_OR
Pahsir	Pahsimeroi - ID	NE_Valleys
Red/Amer	Red/American - ID	N_Cent_Canyons
Salrsf	Salmon R South Fk - ID	Cent_Mts
Sawtrp	Sawtooth Trap - ID	Cent_Mts
Secesr	Secesh - ID	Cent_Mts

Table A-12: Mean survival from tagging to LGR, 1992-1998.

Site	1992	1993	1994	1995	1996	1997	1998
Cathec	0.18	0.23	0.21	0.31	0.24	0.22	0.19
Cfctrp	0.32	0.3	0.19	0.3	0.25	0.53	0.32
Crotrp	0.41	0.26	0.13	0.09	-	0.27	0.23
Grandr	0.3	0.2	0.18	-	0.16	0.26	0.2
Imnahr	0.14	0.22	0.16	0.28	0.28	0.47	0.3
Lemhir	0.24	0.25	0.34	0.42	0.48	0.52	0.38
Loloc	0.3	0.27	0.22	0.04	-	0.44	0.18
Lookgc	-	0.23	0.14	0.24	0.25	0.29	0.26
Lostir	0.25	0.24	0.22	0.22	0.27	0.39	0.31
Marshc	0.14	0.3	0.21	0.37	-	0.57	0.31
Minamr	0.2	0.3	0.15	0.2	0.22	0.23	0.18
Pahsir	0.15	0.24	0.26	0.33	0.32	0.36	0.35
Red/Amer	0.15	0.29	0.13	0.27	0.28	0.33	0.15
Salrsf	0.31	0.19	0.11	0.16	0.15	0.26	0.15
Sawtrp	0.11	0.11	0.18	0.34	-	0.35	0.29
Secesr	-	0.12	0.13	0.13	0.23	0.32	0.24
<i>Annual Average</i>	<i>0.23</i>	<i>0.23</i>	<i>0.19</i>	<i>0.25</i>	<i>0.26</i>	<i>0.36</i>	<i>0.25</i>

Table A-13: Annual average length of fish tagged and annual Palmer Drought Index (PDSI), for 16 sites with 6 - 7 years of tagging data, 1992-1998.

Site	Length at Tagging, mm.							PDSI, July – December, in year of tagging						
	1992	1993	1994	1995	1996	1997	1998	1992	1993	1994	1995	1996	1997	1998
Cathec	77	80	77	87	87	83	79	-3.70	0.22	-2.41	1.60	1.28	-0.16	2.09
Cfctrp	82	77	70	83	82	84	76	-2.30	0.23	-0.95	4.91	4.30	4.34	3.17
Crotrp	82	83	71	77	-	81	78	-2.30	0.23	-0.95	4.91	-	4.34	3.17
Grandr	75	68	71	-	92	80	79	-3.70	0.22	-2.41	-	1.28	-0.16	2.09
Imnahr	73	83	72	84	89	89	87	-2.30	0.23	-0.95	4.91	4.30	4.34	3.17
Lemhir	127	116	108	114	110	112	104	-1.13	1.00	-1.34	5.85	5.40	5.37	5.66
Loloc	76	82	75	109	-	86	68	-2.30	0.23	-0.95	4.91	-	4.34	3.17
Lookgc	-	86	77	91	92	87	86	-	0.02	-1.18	3.47	4.20	2.38	0.12
Lostir	84	72	72	69	88	96	84	-3.70	0.22	-2.41	1.60	1.28	-0.16	2.09
Marshc	71	83	77	93	-	87	74	-3.50	0.06	-3.56	2.33	-	1.18	1.70
Minamr	82	77	68	81	92	76	75	-3.70	0.22	-2.41	1.60	1.28	-0.16	2.09
Pahsir	99	105	96	113	113	113	101	-1.13	1.00	-1.34	5.85	5.40	5.37	5.66
Red/Amer	74	88	68	81	88	79	72	-2.30	0.23	-0.95	4.91	4.30	4.34	3.17
Salrsf	86	72	63	66	69	67	65	-3.50	0.06	-3.56	2.33	1.59	1.18	1.70
Sawtrp	86	-	85	96	-	91	91	-3.50	-	-3.56	2.33	-	1.18	1.70
Secesr	-	61	63	65	70	71	71	-	0.06	-3.56	2.33	1.59	1.18	1.70

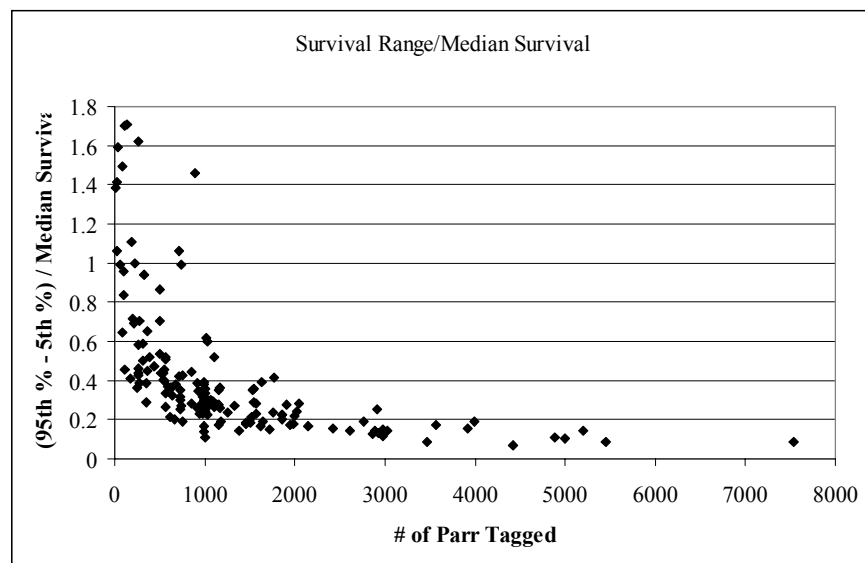
**Figure A-1:** Survival Range/ Median Survival, 32 Tagging Sites, 1992-1998.

Table A-14: 5% and 95% naïve bootstrap confidence limits on survival from tagging to LGR, 1992-1998.

	1992		1993		1994		1995		1996		1997		1998	
	5th %	95th %	5th %	95th %	5th %	95th %	5th %	95th %	5th %	95th %	5th %	95th %	5th %	95th %
Cathec	0.16	0.21	0.19	0.27	0.19	0.23	0.27	0.35	0.20	0.29	0.19	0.25	0.17	0.22
Cfctrp	0.28	0.37	0.27	0.34	0.18	0.21	0.24	0.38	0.20	0.31	0.49	0.57	0.30	0.35
Crotrp	0.21	0.76	0.22	0.32	0.11	0.16	0.03	0.17	-	-	0.19	0.36	0.16	0.32
Grandr	0.25	0.37	0.17	0.23	0.16	0.19	-	-	0.06	0.28	0.23	0.29	0.18	0.22
Imnahr	0.12	0.17	0.20	0.24	0.14	0.18	0.26	0.30	0.26	0.31	0.46	0.49	0.28	0.32
Lemhir	0.21	0.27	0.22	0.28	0.31	0.36	0.34	0.50	0.39	0.60	0.47	0.57	0.36	0.39
Loloc	0.26	0.34	0.25	0.30	0.20	0.24	0.01	0.09	-	-	0.40	0.49	0.16	0.20
Lookgc	-	-	0.21	0.25	0.13	0.15	0.21	0.27	0.09	0.44	0.27	0.32	0.24	0.28
Lostir	0.22	0.29	0.20	0.30	0.19	0.25	0.19	0.26	0.24	0.31	0.36	0.43	0.27	0.35
Marshe	0.11	0.16	0.28	0.31	0.20	0.22	0.31	0.45	-	-	0.54	0.60	0.29	0.32
Minamr	0.17	0.24	0.26	0.34	0.13	0.18	0.17	0.24	0.18	0.27	0.20	0.26	0.16	0.20
Pahsir	0.13	0.17	0.20	0.29	0.24	0.27	0.26	0.40	0.20	0.46	0.30	0.43	0.32	0.38
Red/Amer	0.12	0.18	0.26	0.33	0.12	0.14	0.23	0.33	0.14	0.44	0.30	0.35	0.13	0.17
Salrsf	0.26	0.36	0.17	0.20	0.10	0.12	0.13	0.19	0.13	0.17	0.24	0.28	0.14	0.16
Sawtrp	0.07	0.17	0.06	0.17	0.15	0.20	0.27	0.41	-	-	0.27	0.43	0.25	0.34
Secesr	-	-	0.10	0.14	0.11	0.15	0.10	0.16	0.18	0.28	0.29	0.35	0.22	0.26
<i>Annual Average</i>	<i>0.18</i>	<i>0.29</i>	<i>0.20</i>	<i>0.27</i>	<i>0.17</i>	<i>0.20</i>	<i>0.20</i>	<i>0.30</i>	<i>0.19</i>	<i>0.35</i>	<i>0.33</i>	<i>0.40</i>	<i>0.23</i>	<i>0.28</i>

General considerations for an adaptive management scheme include:

- Probably start small with modest number of sites treated (2-3?), to check for adverse effects.
- Perhaps add 2-3 sites/year, and treat every year after supplementation begins.
- Monitor parr-smolt survival, increasing size of tagged samples to increase precision of survival estimates to 2,000 per site and year, where feasible.
- Add additional treated (nutrient-enhanced) sites over time, assuming no apparent adverse effects.
- Nutrient-equivalent of 1950's spawner numbers (perhaps 1,000 carcasses/site) might provide an upper bound on supplementation inputs.
- Potential threshold effects: Bilby et al 1998 used 0.5-0.7 kg/m² of carcasses. Don't know what this translates into for N, P, but would likely translate to substantial numbers of carcasses.
- Parr-smolt survival could be assessed starting 2nd year of study, recruitment would obviously take longer.
- Smolt-to-adult return rates (SAR's) contrasts between treated and untreated sites are not feasible. One would need to tag more parr than exist in the study areas (on the order of 1 million+).

A.4.2 Monitoring approach

Variables to Monitor

The power analysis (see Section A.4.2.2) focuses exclusively on parr-smolt survival. Other aspects that one might want to monitor include:

- Size of parr at tagging. The larger the parr size, the more likely it is that parr will survive to LGR (see next section), and perhaps survive to adults. Bilby (1998) shows a positive relationship between spawner abundance and parr size for coho and steelhead.
- Number of spawners returning. If supplementation works (i.e., increases survival to adult) this should be higher for treated sites. Increased spawner density may confound the treatment effects: if spawners increase substantially, the need (if any) for additional nutrients may decrease.
- Recruits/spawner. Same rationale as above. *Only* this variable directly addresses extra mortality, *if and only if* one can measure and control for other variables (i.e., spawner abundance, in-river survival, proportion transported, harvest, maturation timing, and upstream survival) that effect the ratio. It is possible that one might be able to test for R/S differences between treated and non-treated sites, but we do not analyze this here.
- Nutrient levels in streams. Supplementation should obviously increase these numbers relative to control sites. Since many spawning/rearing areas are in remote, high-elevation sites monitoring will probably be limited to a few occasions per year, when access roads are snow-free.
- Juvenile densities in rearing areas. These would be expected to increase in treated areas. In addition, increased numbers of juveniles may require more nutrients. Past monitoring of parr density has not produced consistent data that are readily amenable to quantitative modeling (William Thompson, USDA Forest Service, in prep).
- Marine-derived N and P in parr and other stream biota (see Bilby 1999).

Duration and intensity of monitoring

This section discusses the methods and results of a simple power analysis on parr-smolt survival. We first discuss how the data were derived from PIT-tag release-recapture information, bootstrapped to estimate moments of the survival distributions. Next, we demonstrate how this was combined with regional climate data, and used to estimate a simple “base case” model to explain how survival varies among sites and years. Finally, we show how this was used to simulate the power of future experiments to detect changes in survival as a result of (assumed) effects of nutrient supplementation, under a range of experimental designs (# of sites treated, # of years of monitoring, etc).

Creation of data for “base case” model:

1. Extract all wild spring/summer chinook tagged in June-December in the Snake above LGR (approximately 300K tagging records).
2. Extract from (1) records for 16 site in Table A-11, 1992-98, inclusive. Eliminate records with questionable tagging locations or times. Result: approximately 147K records.
3. Bootstrap from (2) 5000 times (with replacement), to obtain 5k survival estimates for each site and year.
4. Use (3) to estimate mean, median, CV, 5th and 95th percentiles of survival data for each site and year.

Base-Case model:

1. Dependent variable is median survival for each site and year (median \equiv mean). Weight for each observation is $1/CV^2$, following Smith (1999).
2. Possible independent variables include mean length at tagging, distance from tagging site to LGR, month of tagging, year effect (dummy variables), site (dummy variables), climate region (dummies), and Palmer Drought Severity Index (PDSI), for various periods before and after tagging.
3. “Best” model, or at least a reasonably good one, includes climate region, year of tagging, length at tagging (for each site and year), and the July-December PDSI, in year of tagging.
4. Base case model equation is Median Survival = intercept + **Region** + **Year of tagging** + Length + PDSI + error, with **Region** and **Year** being dummies.

Results for the base-case are shown in Table A-15. The model explains about 78% of the variation in median survival over the 16 sites and seven years of data, using 13 independent variables for the 105 observations (one per site and year, with a few missing due to lack of tagging data). Cook’s distance diagnostics reveal only one problematic observation – Lolo Creek for 1995. We suspect this is because the survival for that site and year is anomalously low – about 4% -- and not explained well by the model. However, it has little effect on the estimated parameters because it’s CV is quite high. The reported results all include this observation. The Lemhi and Pahsimeroi parr are rather larger than those tagged at other sites, but these observations do not appear to be influential.

Table A-15: Regression results, base case, weighted by $1/(\text{Survival Coefficient of Variation})$. Dependent variable is median survival from tagging to LGR.

Base-Case Regression Results						
Parameter		DF	Estimate	Std Err	Chi Square	Pr>Chi
Intercept		1	-0.3331	0.0726	21.0555	0.0001
Climate Region:	Blues	1	-0.0652	0.0247	6.975	0.0083
	Cent_Mts	1	0.0491	0.0194	6.4252	0.0113
	NE_OR	1	-0.004	0.0217	0.0332	0.8554
	NE_Valleys	1	-0.1371	0.0284	23.341	0.0001
	N_Cent_Canyons	0				
Year of Tagging:	92	1	0.0439	0.0446	0.9687	0.325
	93	1	0.0066	0.0226	0.085	0.7707
	94	1	0.0473	0.0362	1.7045	0.1917
	95	1	-0.0506	0.025	4.0889	0.0432
	96	1	-0.061	0.0295	4.2714	0.0388
	97	1	0.0986	0.0156	40.0862	0.0001
	98	0				
Mean Length at Tagging		1	0.0069	0.0008	70.6881	0.0001
PDSI, July-December		1	0.0201	0.0065	9.46	0.0021
R-Square:	0.782					

Bootstrapping power tests: recall that we want to test the power of detecting a change in survival across a range of years post-treatment, number of sites treated, mean size of the treatment effect, and variation in the size of the effect (e.g., fixed size or drawn from a distribution). Therefore, we did the following:

1. Draw a base-case set of results at random from the 5k sets created in “base-case” data, step 3. Call this set “I”. It will have 7 years of data for each of the 16 sites, again with a few missing.
2. Draw a “post-treatment” set of results at random from the 5k sets. Call this set “J”, with $I < J$. This set will have from 1-7 years of data for 16 sites. The number of treated sites may vary from 1-15, with control sites numbers equal to 16 minus the number of treated sites. Treatment and control sites are assigned at random from among the 16 base-case sites.
3. Add a treatment effect to each year of simulated survival data for each treated site selected in (2), above. This effect may be either fixed or drawn from a normal distribution. Note that other than survival at treated sites, the expected value for all variables in the post-treatment set “J” is the same as for the base-case set “I”. However, both survival and length at tagging will differ between the two, since they are drawn from two different outcomes of the 5k bootstrap games created previously.
4. Estimate a model identical to the base-case model previously described, but with a “treatment” dummy variable for the treated sites. If this treatment effect is significant at 0.05, the game is assigned a “1”, otherwise it is assigned a “0”.
5. Power is measured as the proportion of tests that have a “1”, for the # of power-test games are performed.

The above simulations were repeated from 100-1000 times. The power results appear to converge reasonably well after 100 or so iterations, but there are a few anomalies that don’t affect the conclusions.

Note that several assumptions are implicit in this procedure. First, treatment (nutrient or carcass supplementation) is assumed to be in effect for each treatment site for each year post-treatment. Second, the independent variables other than length at tagging (region, the tagging year effects, and PDSI) are identical pre- and post-treatment. Length changes only because it was estimated separately for each of the 5k base-case games. This amounts to assuming that climate, at least as measured by the PDSI, can be represented post-experiment by the pre-experiment years of data. Although we have about 90 years of PDSI data available, we have not yet tested this assumption.

Results are shown in Tables A-16 – A-18. The results can be interpreted as follows, using the 1st row of Table A-16 as an example. For an effect size of 0.025, 1 year of the experiment, and 1 treated site, 8% of the power tests were significant at 0.05 or better. As the effect size increase through 5%, 7.5%, and 10%, again with 1 treated site and 1 year of post-treatment data, the proportion detected correctly (at 0.05 or better) increases from 8% to 12%, 22%, and 32%. Looking at the 4th-to-last row (7 years post-treatment, 7 treated sites), the power increases from 48% to 89%, 99%, and 100% for effect sizes of 2.5%, 5%, 7.5%, and 10%, respectively.

Table A-16 shows results with effect sizes “fixed,” or assuming no variation in effect size. As one would expect, power increases with effect size (i.e., difference between survival with and without treatment), and with the number of years post-treatment. In addition, within a given number of years post-treatment and effect size, power usually increases with the number of sites treated, up to about 7 or 9, and then decreases slowly as the proportion of treated sites increases to more than half of the 16 sites. For some reason (we’re not sure why) power is higher for 15 treated sites (i.e., only one control site) than when treating only a single site, with 15 controls.

Table A-16: Power of ability to detect additive survival increase, assuming no variation in treatment effect.

Years Post-treatment	# of sites treated (of 16)	True effect = 0.025, Power (Proportion detected "correctly") @ 0.05	True effect = 0.050, Power (Proportion detected "correctly") @ 0.05	True effect = 0.075, Power (Proportion detected "correctly") @ 0.05	True effect = 0.10, Power (Proportion detected "correctly") @ 0.05
1	1	0.08	0.12	0.22	0.32
1	3	0.1	0.38	0.52	0.63
1	5	0.2	0.32	0.56	0.71
1	7	0.14	0.48	0.72	0.82
1	9	0.24	0.51	0.66	0.81
1	11	0.18	0.45	0.69	0.85
1	13	0.14	0.41	0.5	0.6
1	15	0.14	0.32	0.37	0.56
3	1	0.06	0.26	0.43	0.54
3	3	0.3	0.54	0.73	0.87
3	5	0.29	0.68	0.93	0.98
3	7	0.33	0.73	0.93	1
3	9	0.33	0.71	0.94	0.99
3	11	0.37	0.68	0.93	0.97
3	13	0.38	0.69	0.86	0.91
3	15	0.26	0.5	0.7	0.82
5	1	0.12	0.27	0.41	0.64
5	3	0.34	0.66	0.75	0.91
5	5	0.38	0.68	0.95	0.98
5	7	0.5	0.78	0.96	1
5	9	0.39	0.79	0.97	1
5	11	0.4	0.81	0.95	1
5	13	0.41	0.7	0.92	0.97
5	15	0.38	0.55	0.82	0.94
7	1	0.18	0.28	0.61	0.52
7	3	0.23	0.55	0.83	0.97
7	5	0.43	0.81	0.95	0.99
7	7	0.48	0.86	0.99	1
7	9	0.46	0.81	0.99	1
7	11	0.52	0.78	0.98	1
7	13	0.46	0.73	0.92	1
7	15	0.35	0.6	0.82	0.91

Tables A-17 and A-18 display results for the most powerful type of tests – 7 treatment sites (and 9 controls), for 7 years post-treatment, with effect sizes drawn from normal distributions of different means and variances, as shown. In some respects, the results are more or less what one would expect: as variance in effect size increases, power decreases, all else held equal. However, for effect sizes that are reasonably powerful at low variance (5% and above), the variance can increase markedly without decreasing power by too much. This trend is continued in Table A-18: the mean effect can be much smaller than it's standard deviation (see last few rows) without decreasing power dramatically. We believe a partial explanation is the normal distribution assumed for the effect size: for every anomalously small value of the effect, an anomalously large one will also be drawn, and the two balance one another.

Table A-17: Power of ability to detect additive survival increase, with variation as noted in treatment effect. All run with seven treatment sites and seven years post-treatment.

Variance	Std. Dev.	Power, Effect size = 0.01	Power, Effect size = 0.03	Power, Effect size = 0.05	Power, Effect size = 0.07	Power, Effect size = 0.09	Power, Effect size = 0.11
0.001	0.03	0.3	0.61	0.84	0.98	0.98	1
0.003	0.05	0.23	0.61	0.74	0.97	1	1
0.005	0.07	0.19	0.54	0.76	0.96	1	1
0.007	0.08	0.22	0.53	0.72	0.86	0.99	0.99
0.009	0.09	0.27	0.47	0.71	0.88	0.96	0.99
0.011	0.10	0.23	0.49	0.76	0.89	0.97	0.99
0.013	0.11	0.22	0.43	0.71	0.87	0.94	0.97
0.015	0.12	0.32	0.43	0.58	0.86	0.93	0.98
0.017	0.13	0.29	0.47	0.6	0.8	0.94	0.98
0.019	0.14	0.23	0.48	0.72	0.81	0.91	0.95
0.021	0.14	0.24	0.37	0.65	0.7	0.88	0.97

Table A-18: Power of ability to detect additive survival increase, with (more) variation as noted in treatment effect. All run with 7 treatment sites and 7 years post-treatment.

Variance in Effect Size	Std. Dev. of Effect Size	Power, Effect Size = 0.05	Power, Effect Size = 0.07	Power, Effect Size = 0.09	Power, Effect Size = 0.11
0.023	0.15	0.58	0.79	0.91	0.96
0.025	0.16	0.60	0.74	0.90	0.93
0.027	0.16	0.60	0.74	0.86	0.93
0.029	0.17	0.58	0.76	0.87	0.95
0.031	0.18	0.57	0.73	0.84	0.92
0.033	0.18	0.57	0.71	0.84	0.94
0.035	0.19	0.55	0.68	0.80	0.91
0.037	0.19	0.54	0.71	0.79	0.90
0.039	0.20	0.57	0.69	0.81	0.91
0.041	0.20	0.53	0.64	0.79	0.90
0.043	0.21	0.53	0.67	0.80	0.89
0.045	0.21	0.51	0.70	0.78	0.88
0.047	0.22	0.52	0.65	0.79	0.88
0.049	0.22	0.50	0.66	0.78	0.90
0.051	0.23	0.49	0.65	0.76	0.86
0.053	0.23	0.49	0.65	0.79	0.84

A.4.3 Benefits, risks, costs, and trade-offs

Benefits and Amount of Learning Possible

We need to work more on metrics here. Without some way of comparing trade-offs in extinction risks, experiment/monitoring costs, and other factors, it's going to be difficult to communicate this well to audiences outside PATH.

Risks to Stocks

- Disease spread is possible if carcasses are used.
- “Surprises” (both pleasant and unpleasant ones) obviously possible.

Costs

Obvious ones are:

- Increase in tagging effort (cost of tags and field researcher time).
- Fertilizer purchase and application.
- Time needed for carcass outplanting (assume cost of carcasses = 0).
- Spawner #'s and age may need to be monitored in areas where this is not done at present.

Trade-offs

A.4.4 Inferences

Anything beyond the obvious [i.e., smolt or R/S survival as $f(\text{fertilization})$]?

Table A-19: Observations and inferences for nutrient-driven stock viability hypothesis.

Variable	Observation and Inference	
	Observations Consistent with Nutrient-Driven Stock Viability Hypothesis	Observations Not Consistent with Nutrient-Driven Stock Viability Hypothesis
Parr-smolt survival	Increase (in fertilized streams), relative to controls	Decrease or no change (in fertilized streams), relative to controls
Parr Size	Increase (in fertilized streams), relative to controls. Assumes that fertilization effects egg-parr growth rates.	Decrease or no change (in fertilized streams), relative to controls
V_n	N/A, unless “enhanced” smolts perform differently	N/A, unless “enhanced” smolts perform differently
Spawner #'s and ages for S/S survival	Increase (in fertilized streams), relative to controls	Decrease or no change (in fertilized streams), relative to controls
R/S	Increase (in fertilized streams), relative to controls	Decrease or no change (in fertilized streams), relative to controls

A.4.5 Confounding factors

Good design should be able to avoid most confounding, since real controls appear to be possible. One possible confounding factor is smolt or parr density and its effects on survival.

A.4.6 Practical constraints

- The number of extra carcasses available may be a limitation.
- Public support would be needed, especially for actions on privately owned land.

A.5 Manipulate Hatchery Production

A.5.1 Description of experimental action/research & monitoring

Rationale

Study Objective: To determine if: 1) there is support for the stock viability extra mortality hypothesis (i.e., that something unrelated or additional to hydrosystem development has accounted for the total “extra mortality” [including D] estimated since the mid-1970s); and 2) reducing or eliminating exposure of wild Snake River spring/summer chinook migrants to hatchery steelhead can reduce total “extra mortality” of spring/summer chinook in the future, without breaching four Snake River dams. By simultaneously monitoring variables used to estimate D (Section 3.1), and/or by simultaneously conducting transportation experiments (Sections 3.2-3.3), relative impacts of hatchery steelhead production on transported vs. non-transported spring/summer chinook can be estimated.

Description of Hypothesis: Rationale for the stock viability extra mortality hypothesis and, specifically, the assumption that hatchery steelhead production is a causal factor, has been reviewed in the PATH August 1998 Weight of Evidence report and supporting documents. Briefly, Snake River hatchery smolt production increased greatly following the construction of the Lower Snake River dams. Steelhead production in particular increased from approximately 4 million smolts released per year to approximately 10 million smolts per year during the 1980’s. The increase in hatchery production in the Snake basin coincides with increases in ‘extra mortality’ (including D) estimated for Snake River spring/summer chinook (e.g., Williams et al. PATH WOE Submission #1, 1998). Possible mechanisms for a negative effect of hatchery fish on co-mingled wild spring/summer chinook juveniles include: 1) delayed mortality resulting from stress of exposure during the outmigration from the upper Snake to below Bonneville Dam; 2) delayed mortality resulting from stress induced by interactions during periods of delay at hydropower projects or in the barge/collection systems; and 3) negative interactions in the lower river/estuary exacerbated by the relatively poor condition of wild Snake River spring/summer chinook migrants.

It is possible that any negative effects of hatchery production on wild Snake River spring/summer migrants are a result of a combination with hydropower effects. In that case, changes to the hydropower system may relieve mortality due to hatchery interactions. For example, effect (2) might be exacerbated by the lack of effective separation of hatchery steelhead from yearling chinook prior to holding in raceways and loading on barges. No separation occurs at Lower Granite Dam and separation efficiency at other collection projects ranges from only 36-71%. Future improvements in separation efficiency might eliminate at least part of the extra mortality, without reducing numbers or size of hatchery steelhead. However, it is also possible that hypothesized negative impacts of hatchery production may not be relieved by changes to the hydropower system. In that case, changes in the hatchery program would be necessary to relieve mortality effects on wild Snake River spring/summer chinook migrants.

Arguments against the stock viability hypothesis and, specifically, against the possibility of hatchery steelhead production as a causal factor, are summarized in the PATH August 1998 Weight of Evidence report. An experimental approach to evaluating the effect of hatchery production on “extra mortality” would attempt to resolve the differing interpretations of currently available information.

Experimental Action: Manipulate Snake River hatchery steelhead production to reduce exposure of wild Snake River spring/summer chinook juveniles to relative levels at or below those experienced in the 1970’s. Hatchery steelhead exposure with wild Snake River spring/summer chinook juveniles could be reduced in several ways including reducing the number of steelhead smolts released, reducing the size of steelhead smolts at release (reducing steelhead biomass), or delaying steelhead smolt releases until late in the migration season. It would be desirable to alternate or vary relative exposure across a series of brood years, taking advantage of the contrast to evaluate the relative effect of reduced exposure.

The exact experimental design would need to be developed as part of subsequent PATH experimental management tasks. However, two alternatives have been explored to provide examples of the efficacy of possible approaches. Each starts with a quantification of the hypothesized effect of hatchery steelhead production on total extra mortality (including D). A simple linear relationship between (m-M) and SH hatchery releases (as derived from WOE Submission 1, Figure 5) yields the functional relationship shown in Figure A-2.

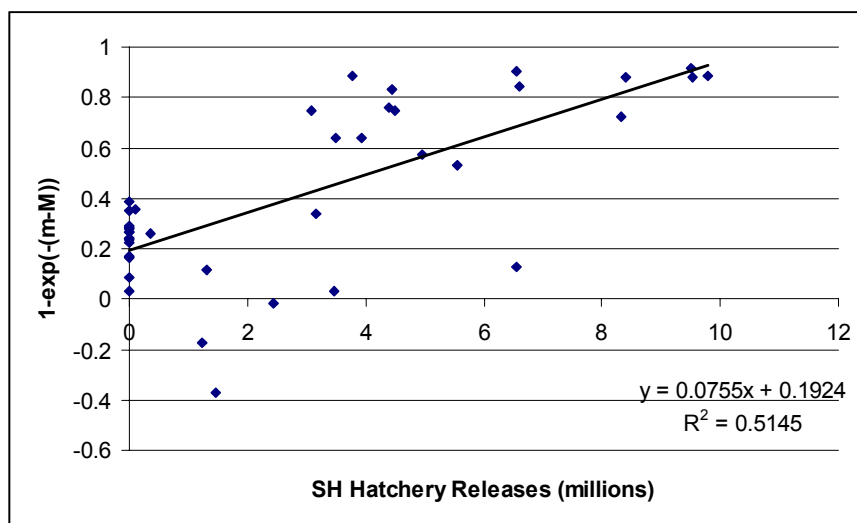


Figure A-2: Regression of Snake River spring/summer chinook total extra mortality (including D), as determined from the PATH Delta model, and steelhead hatchery releases affecting 1952-1992 brood years.

Spatial and Temporal Components

Spatial contrasts are not possible with this approach because hatchery interactions are hypothesized to occur throughout the mainstem Snake and Columbia River. Temporal contrast would be generated by curtailing hatchery production or shifting release levels forward in time during treatment years to reduce exposure of wild migrating spring/summer chinook to hatchery steelhead. Treatments could either be in alternating years or in alternating brood cycles. The objective would be to reduce exposure in treatment years to no higher than the levels experienced in the late 1970s.

Two hypothetical experiments illustrate possibilities for generating temporal contrast in treatments (Table A-20). In these examples, the treatments are held constant for 5-year intervals approximating brood cycles. Hypothetical Experiment 1 would attempt to generate extreme contrast by increasing hatchery releases well above current levels to 20 million smolts in one treatment and by reducing hatchery production to 1 million smolts in the alternating treatment. Note that this example is provided only to show the effects of a somewhat extreme degree of contrast among treatments. If one actually wanted to implement this experiment, new production facilities would have to be built to produce the 20 million smolts. Other practical implications are discussed below. Hypothetical Experiment 2 would compare hatchery releases near current levels (8 million smolts) to a level more similar to that in the 1970s (4 million smolts).

Table A-20: Hypothetical examples of two possible experiments to evaluate effects of hatchery steelhead production on Snake River spring/summer chinook salmon survival. Experiment 1 represents a high-contrast, five-replicate experiment, while Experiment 2 represents lower contrast among treatments and only two replications.

Experiment	Minimum smolt releases	Maximum smolt releases	Interval	Duration	Start of experiment
1	1 million	20 million	5 years	50 years	Year 10
2	4 million	8 million	5 years	20 years	Year 10

A.5.2 Monitoring approach

Smolt-to-adult returns (SAR) and returns per spawner (R/S, or the difference between R/S and predicted R/S = RES) of the wild spring/summer chinook index stocks would be the primary response variables. Survival of fish from the alternating treatment periods would be compared to determine if there is an effect of hatchery releases. Lower river index stocks would need to be monitored to account for common year effects that could fortuitously coincide with different treatment periods. In-river survival (V_n) and ratios of transported and non-transported SARs would need to be monitored concurrently (Sections 1.1-1.3) to draw inferences about the relative effects of changes in hatchery releases on D and extra mortality of in-river migrants. See Table A-21 for a summary. If possible, data from PIT tagged groups would be used in the analyses.

Duration and Intensity of Monitoring

The expected duration of the experiment would depend largely upon the details of implementation of the reductions in exposure to hatchery steelhead. For the two hypothetical alternatives described in Table A-20, an analysis by C. Peters (July 26, 1999 report) suggests that treatment effects would be discernable with the high-contrast, long- duration (50+ years), Experiment 1 (Figure A-3). However, treatment effects are not likely to be seen with the lower-contrast, shorter duration (20 years) Experiment 2 (Figure A-4). Future analyses would be necessary to explore additional experimental options that are intermediate to the two examples presented in this report.

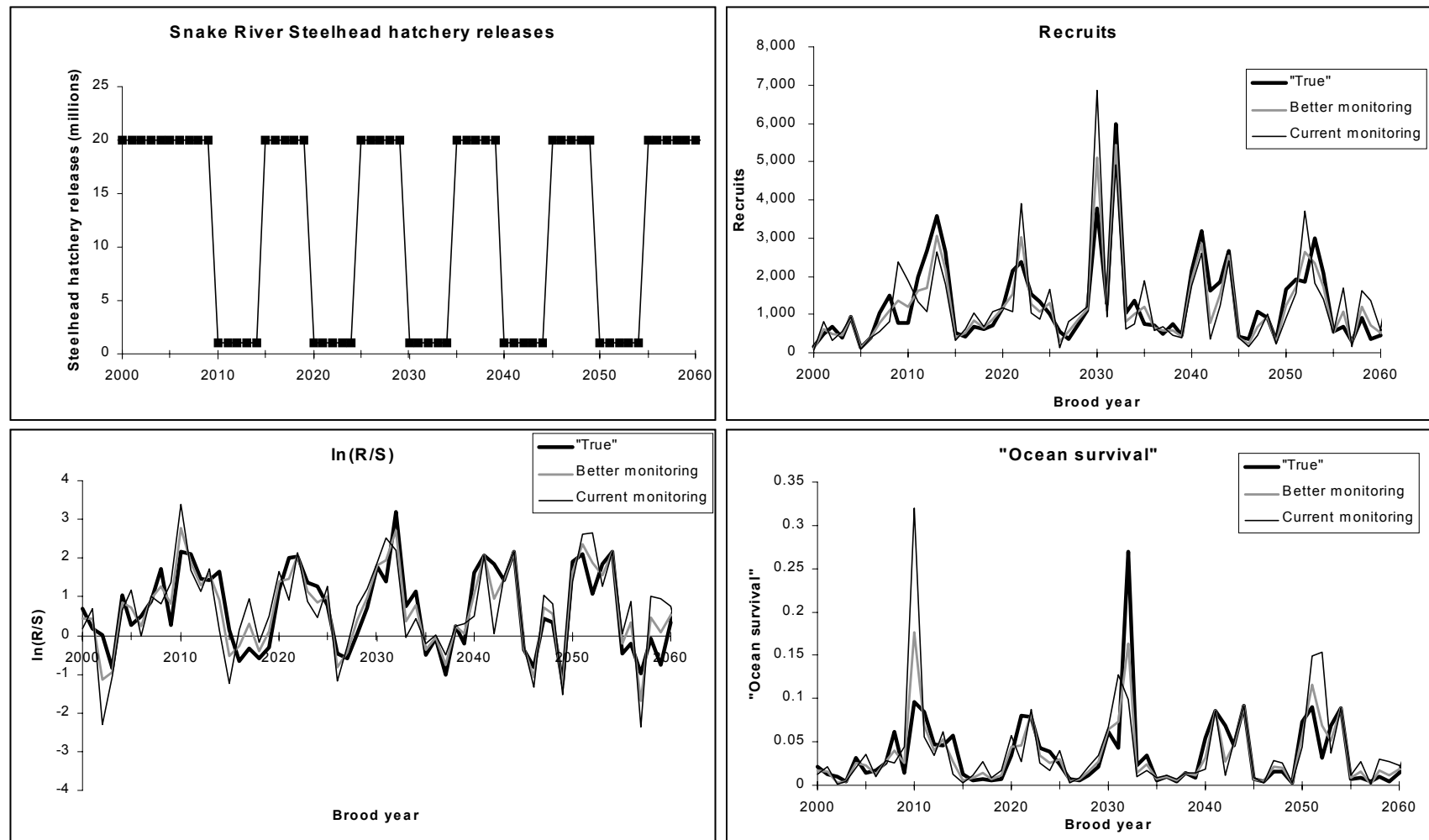


Figure A-3: Simulation of expected results for hatchery Experiment 1 from Peters (1999). "Better monitoring" refers to a sensitivity to reducing random error in simulations to 50% of base value.

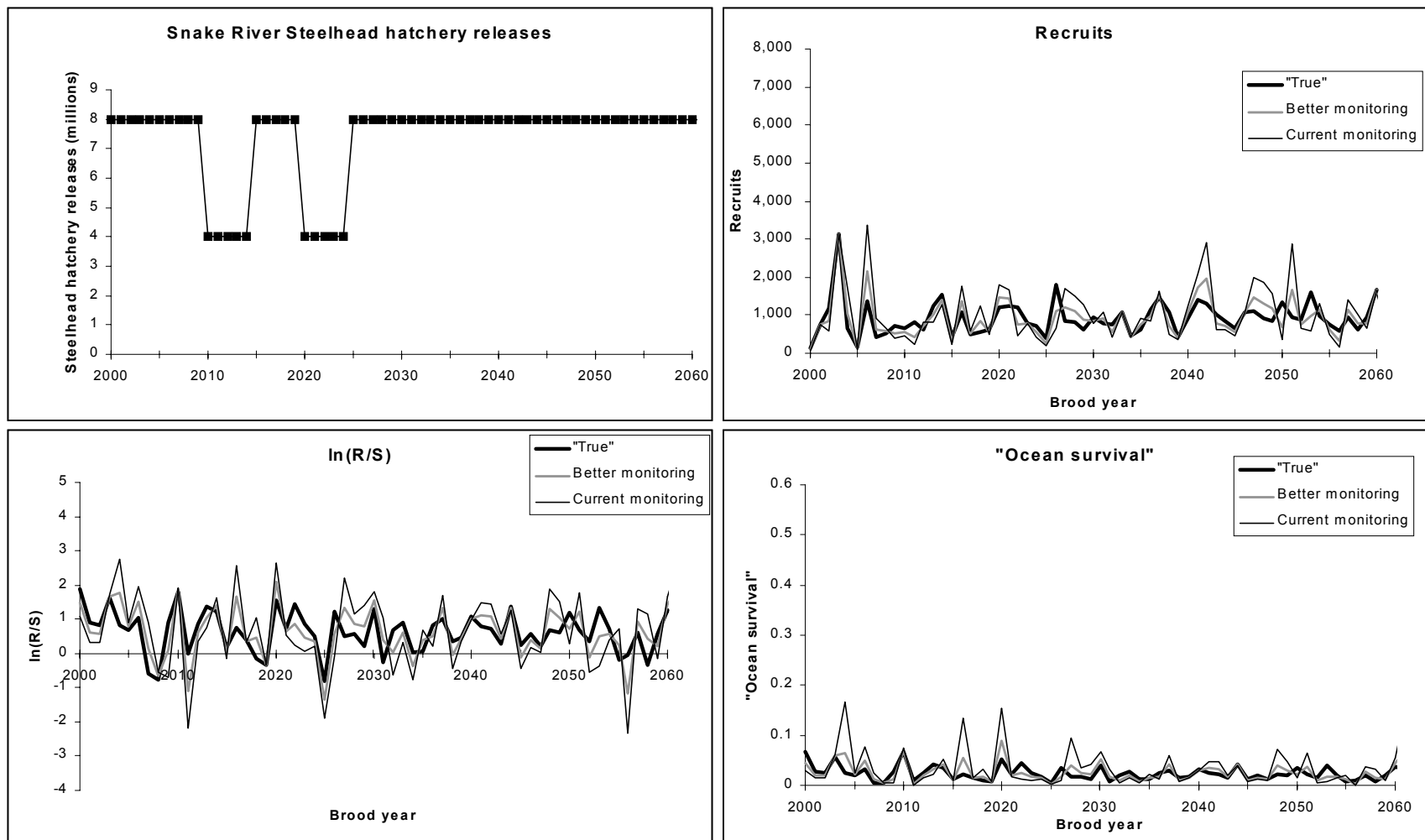


Figure A-4: Simulation of expected results for hatchery Experiment 2 from Peters (1999). "Better monitoring" refers to a sensitivity to reducing random error in simulations to 50% of base value.

A.5.3 Benefits, risks, costs and trade-offs

Benefits and Amount of Learning Possible

As has been noted in the review of the Weight of Evidence Report by the SRP, the effects of alternative causative factors on extra mortality of spring/summer chinook are confounded in time. Evaluating the contribution of increased hatchery production to extra mortality of spring/summer chinook would clarify the long-term response of these populations to alternative management actions. The results of such a study could help determine which combinations of hydropower actions and hatchery management scenarios are most likely to result in achieving recovery goals for Snake River spring/summer chinook. Given the possibility that increased hatchery production has contributed to increased mortality of outmigrating spring/summer chinook, pairing experimental hatchery studies with whichever hydropower strategy is chosen would allow managers to adjust programs in the future to achieve rebuilding goals.

Risks to Stocks

The goal and cap for hatchery production of all stocks, as part of the Lower Snake Compensation Plan, is 20 million. The goal for steelhead is about 14 million. However, current programs are producing about 12+ million smolts per year because of limits imposed by ESA biological opinions and current fish culture practices. Because the production goal has rarely been met and because it is the number of fish released (not the production goal) that determines the potential for negative interaction, this paper focuses on potential hatchery reductions in terms of recent year releases (not production). Steelhead are produced at USFWS, ACOE, and Idaho Power facilities.

Steelhead releases in the Snake River in 1998 totaled 12.2 million. Approximately 3 million of these releases were used for conservation and/or restoration. 'Conservation' is defined here as programs based on native or local stocks, and 'local' means part of an ESU. Based on recent release numbers, this leaves a possible maximum reduction in hatchery steelhead releases of 9.22 million from all hatcheries combined without impacting conservation/restoration programs. Options for reducing exposure of spring/summer chinook to steelhead smolts should take into account the desire of maintaining hatchery program broodstocks to allow for rapid return to levels consistent with mitigation responsibilities. Estimating the degree to which reductions could occur before the otherwise successful steelhead hatcheries would become crippled is a tenuous proposition. There would also be some institutional resistance, and possibly legal barriers, to reducing the effectiveness of steelhead hatcheries.

A high-contrast experiment such as Hypothetical Experiment 1 would involve a special set of practical constraints. NMFS estimates that the equivalent of 3-5 new large steelhead hatcheries would be necessary to produce the 20 million smolts needed for the high release treatment. Also, the 20-fold difference in treatments (going from 1 million to 20 million smolts in one brood cycle) would be difficult, if not impossible, to achieve, based on past experiences with the time required to build hatchery production capability.

Costs

Manipulating or reducing a major portion of the current hatchery production of steelhead in the Snake River basin would have substantial costs, both monetarily and in terms of risks to the future hatchery production program. To the extent it is feasible to reduce interactions by delaying releases at major facilities, additional manpower and feeding costs would be incurred. Other production programs at those hatcheries might be negatively affected by the need to allocate rearing space to steelhead for longer period of time. Post-release survival of steelhead smolts may be impacted by delayed releases. If it were necessary to reduce programs for some period of time to implement the experiment, additional costs of

mothballing facilities and programs would be incurred. As described above, if high-contrast experiments required production above current levels, this would also result in significant additional costs.

If we discovered that reductions in hatchery fish were insufficient for recovery or detrimental to conservation efforts and the hatcheries were turned back on, there would be biological limitations and considerable costs. Historically, the length of time it took for the LWSCP hatcheries to meet their production goals varied dramatically across the various hatcheries and depended on the survival rate and return of adults, water temperature limitations, disease problems etc. For example, the first LSRCF facility (Grande Ronde R. basin) was completed in 1978, but the hatchery did not meet its production goal of 1.35 million until 1986, 8 years later. In addition, if hatcheries were turned back on under the condition that local stocks be used for broodstock, it is unlikely that the composition of returning adults would be appropriate to support the broodstock needs of the supplementation program for many years. Costs would include re-hiring or transferring staff back, taking the hatcheries out of moth-ball or maintenance mode, and running hatcheries at minimal production until broodstock could be built back up. Turning hatcheries back on could take from 5 to 10 years, depending on broodstock requirements and those factors listed above.

Tradeoffs

Under a scenario of status quo management of the hydrosystem, where transportation is maximized, reduced hatchery releases may have the potential to increase spring chinook returns. It is extremely important to note, however, that this option would require consensus from all those groups involved in Columbia River Fishery Management Plan renegotiations, may require congressional approval if the Lower Snake Compensation Plan Act and USFWS treaty trust responsibilities are violated, and would likely result in a substantial reduction in state and tribal fisheries. Thus, this scenario has complex management implications that require consideration of fishery regulations and treaty rights and how these may be affected by reductions in hatchery steelhead.

There may be alternatives to conducting the proposed experimental management action. Although limited, some data are available that would allow comparisons of spring chinook survival (SARs) during periods when densities of hatchery fish are high and when they are low. Data presented to date indicate SARs increase during periods of low density but also indicate SARs are still low during these periods (well below two percent). It is possible that additional analyses of existing information would remove the uncertainty regarding the influence of hatchery production on total extra mortality (including D). Some PATH members believe that additional analyses of available data should be conducted to assess the potential benefits of reduced hatchery production to spring chinook prior to implementing reductions that could harm affected hatchery stocks. Others believe that because of the confounding effects described in previous documents (e.g., August 1998 PATH Weight of Evidence report), it will not be possible to resolve this question without some form of management experiment. SRP comments also appear to support the second opinion.

The stock viability hypothesis, with hatcheries as a causal factor, has only been proposed to explain patterns of extra mortality in Snake River spring/summer chinook salmon. There would be little opportunity for improving fall chinook survival as a result of this experiment because there is little overlap in time between outmigration of falls and spring chinook and steelhead. The potential for improving other stocks such as sockeye and coho has not been assessed. *Note: this is also true of most of the other proposed experimental management actions.*

A.5.4 Inferences

Table A-21: Example inference table for Hatchery-Caused Stock Viability Extra Mortality Hypothesis.

Variable	Observation and Inference	
	Observations consistent with “SV-Hatchery”	Observations not consistent with “SV-Hatchery”
V_n	Must be monitored to concurrently estimate D, but not directly relevant to testing extra mortality hypothesis. Same with transport and non-transport SARs.	
SAR or RRS	Higher in reduced hatchery production years than in higher production years.	No difference among treatments, or lower in higher production years
λ_n	If possible to infer from estimate of RES and D; should go down in reduced hatchery production years	No change among treatments or goes up in reduced hatchery production years

A.5.5 Confounding factors

As described previously, it is likely that hydro and hatchery factors that may be responsible for extra mortality are confounded. If changes to the hydrosystem are being made between treatments, these may confound results, especially if there are few replicated treatment blocks.

A.5.6 Practical constraints

These were described in detail in Section A.5.3.

A.6 4-dam Drawdown

A.6.1 Description of experimental action / research & monitoring

Rationale

Description of Hypothesis: The completion of the Federal Columbia River Power System in the late 1960's through the mid-1970's and subsequent operation, has increased the direct and delayed mortality of juvenile migrants, which resulted in considerably sharper declines in survival rates of Snake River spring and summer chinook stocks (over the same period), than of similar stocks which migrate past fewer dams and are not transported.

PATH retrospective analyses (PATH FY96; Conclusion 3a.1) concluded that the differences in stream-type chinook indicators of productivity and survival rates between upstream (Snake) and downstream (Lower Columbia) are coincident in time and space with development of the hydrosystem (high confidence). PATH also concluded that, on a decadal scale, differences in these indicators between Upper Columbia and Lower Columbia are coincident in time and space with development of the hydrosystem (reasonable confidence, low confidence with regard to specific years).

SNAKE River fall chinook also declined following Snake River dam construction and operation, whereas similar stocks above fewer dams (Hanford - 4 dams; Lewis – 0 dams) have remained more stable. Snake River steelhead declines were also temporally associated with Snake River dam construction and operation.

The proposed experimental action recognizes that two major hydropower treatments already have been applied to upriver stocks, construction and operation of dams and juvenile fish transportation. This proposed experimental action partially reverses the treatments for listed Snake stocks (consistent with ESA requirements to ensure survival and recovery), and evaluates the magnitude of the response. Regional stock responses would be used to: (1) determine the extent to which dam removal affects survival and recovery of Snake River stocks; and (2) evaluate likely effects for decisions (John Day drawdown) on the listed Upper Columbia stocks.

Experimental action: Breach Snake River dams, stop transportation, evaluate regional stock responses to help guide John Day drawdown decisions for listed Upper Columbia stocks. Hatchery production could be either pulsed or kept constant under this approach (assumed constant in this option).

Explicit Objectives: Recover listed Snake River salmon and steelhead populations, determine consistency of Snake River population response to alternative hypotheses about delayed or extra mortality, and evaluate hypotheses relevant to future management decisions, specifically for recovery of upper Columbia River listed populations.

The stated purpose of experimental management (Section 1.1) is to “...**both maximize the ability to achieve conservation and recovery objectives, and concurrently learn something about key uncertainties to improve future management.**” This experimental option proposes reductions in direct and delayed mortality of Snake River stocks using the most risk-averse hydropower action to provide a large contrast in stock response for evaluation of mortality components. The magnitude of the observed change would be contrasted with that projected from alternative PATH hypotheses about extra mortality, to evaluate consistency of hypotheses with empirical data, improving the predictive capability for future management decisions, specifically for listed upper Columbia River populations. The timing and sequence of actions are based on earliest feasibility of implementation assumed in previous PATH analyses.

Testable hypothesis: Following Snake River dam breaching (A3/A5), the measured (estimated) values of R/S residuals, μ and relative change in SAR will best fit those projected by the one of following extra mortality hypotheses: (1) Hydropower, (2) Stock Viability, or (3) Regime Shift.

To test this hypothesis, projections of R/S response, μ and differential SAR specific to each regional contrast (Snake, Lower Columbia, Upper Columbia) would first be made (in FY99) using passage/transport models, which produce different ranges of in-river survival, T/C ratios and D-values. For example, non-hydropower extra mortality hypotheses for spring/summer chinook are expected to project a substantially smaller reduction in the Snake River μ (and also less relative change in SAR) following A3/A5 implementation. The projected values are specific to passage models and estimated D-values (from T/C and in-river survival estimates). A pattern of greatly reduced Snake River μ , relative to change in upper Columbia μ , would be evidence for the hydro extra mortality hypothesis.

Statistical hypothesis or decision rule: A framework is presented in Appendix D to relate future monitoring data to PATH life cycle models to help test hypotheses regarding the magnitude of responses to management actions. Measured responses in R/S residuals (i.e. Relative Recruitment Success (RRS)) and differential SARs would be compared to projected responses to determine which hypotheses best fit the data.

To test R/S response for spring/summer chinook, residuals from the R/S data from upstream and downstream stocks are measurable empirically and correspond to terms in the delta model. Now consider the differences in performance between upstream and downstream stocks. We would like to see if an

action changes the performance of upriver stocks relative to downriver stocks. Though it would be nice to know whether an improvement due to some action occurs in system survival or extra mortality, the most important thing to know is that $(RRS_u - RRS_d)$ is positive (i.e., the status of Snake River stocks is improving relative to downriver stocks). To assess the response of the system to implementing a natural river option, for example, we measure total mortality “m” and see how much it changes (see Appendix D for derivation using the delta model). That is,

$$RRS_u - RRS_d = m - \ln(V_d) \quad [\text{Eq. D-9}]$$

Note that equation D-9 is analogous to the parameter ‘ μ ’ referred to throughout this section. This formulation would require an estimate of in-river survival (V_d) of smolts from the lower river tributaries (i.e., John Day River).

A comparison of [b]efore versus [a]fter conditions would attempt to measure the changes in the upriver-downriver differences in the residuals, that is:

$$\{RRS_u - RRS_d\}[a] - \{RRS_u - RRS_d\}[b] = m[a] - m[b] - \ln(V_d[a]/V_d[b]) \quad [\text{Eq. D-10}]$$

Therefore the only thing we need to factor out is changes in the in-river survival of downriver stocks ($V_d[a]/V_d[b]$). Then we can directly measure the net benefit of an action in terms of $m_a - m_b$.

Appendix D also presents equations to address the question of where the net benefits occurred (i.e., improved system survival or post BON survival), but cautions that this becomes more difficult to determine.

We can model SARs in a similar manner. Smolt-to-adult return rates of upriver and downriver stocks would be estimated and contrasted as:

$$\ln(\text{scaled SAR}_u) - \ln(\text{scaled SAR}_d) = m - \ln(V_d) \quad [\text{Eq. D-16}]$$

This is analogous to eq. D-9, and assumes that upriver and downriver stocks have similar ocean mortality. So tagging should in principle be an alternative way to get at total mortality rate, “m”. The SAR data involve fewer unknown coefficients, since the egg-to-smolt survival is not part of the estimate. Therefore, there is one less source of variation. In addition, observed SARs can be directly compared to the PATH goal of 2% to 6% needed for survival and recovery of Snake River spring/summer chinook (FY98 report).

Spatial and temporal components

Experiment period: Experiment period is 8 years (depending somewhat on the definition). There are actually four periods; pre-1970, 1975-2003 (implement A3/A5), 2004-2012 (evaluate effects of A3/A5), post-2012 (implement B1/B2, depending on results of evaluation).

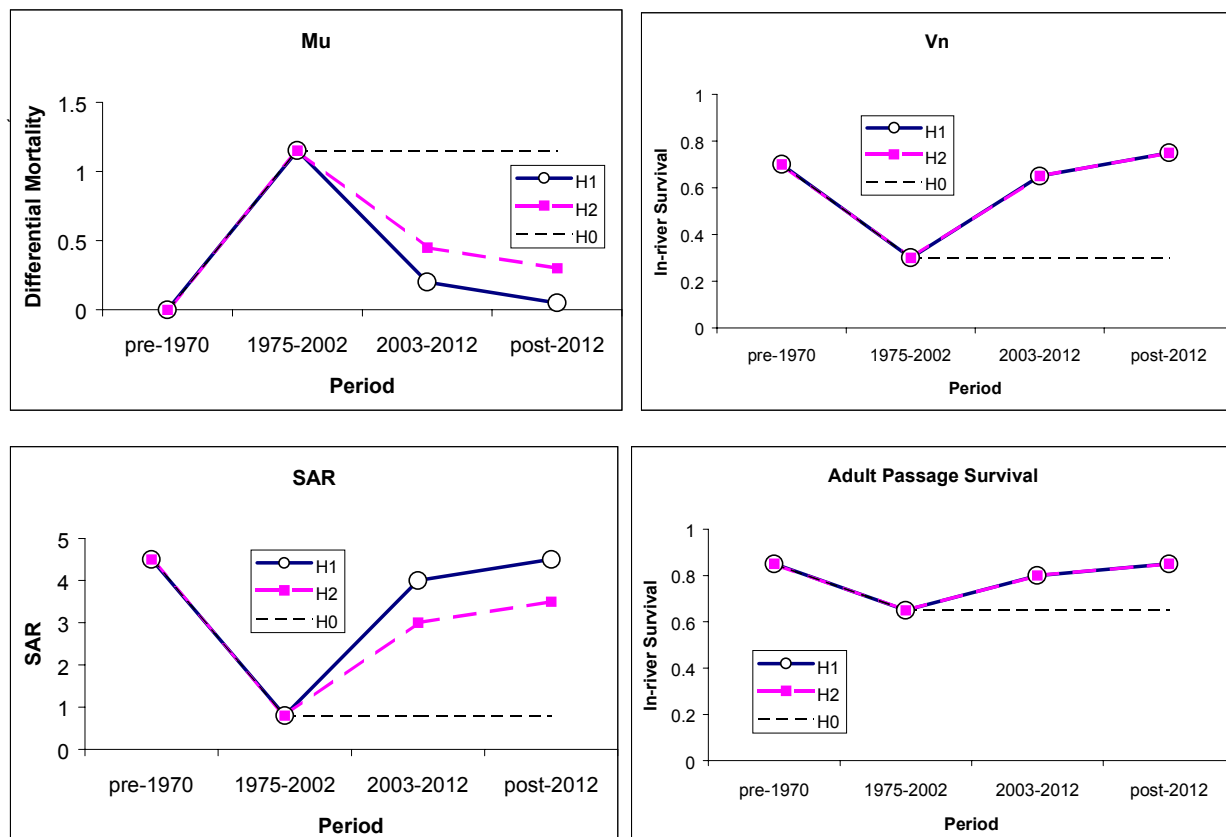
Spatial resolution: Regions and stock groupings of the interior Columbia Basin.

Experimental units: Regional stock groupings (stream-type chinook, ocean-type chinook, steelhead in the Snake, Upper Columbia and Lower Columbia regions). Index stocks are replicates within the regional groupings.

Treatments Interspersed in Time: Implement A3/A5 for Snake River populations (3 region-species groups) in 2003, and B1/B2 for upper Columbia populations (3 region-species groups) in 2012.

Hypothesized response for Snake River stocks would be large reductions in μ for the 2003-2012 period following implementation of A3/A5 (Fig. A.6-1), reflecting decreases in direct and extra (delayed) mortality. Implementation of B1/B2 in 2012 would further reduce μ for Snake River stocks. The expected change in μ depends on the alternative hypotheses about extra mortality and D. The hydro hypothesis was illustrated as H1 and Stock Viability hypothesis as H2 in Figure A.6-1. (note: values are for illustration; to be replaced with PATH results). Similarly, the hypothesized SAR response would be improvements for Snake River stocks beginning in 2003, with an additional increase beginning in 2012 (Fig. A.6-1). H1 and H2 project similar increases in in-river survival rate and upstream passage survival rate following A3/A5

Hypothesized Snake River Response



and B1/B2 implementation (Fig. A.6-1).

Figure A.6-1: Hypothesized change for Snake River stocks in differential mortality (μ), SAR, in-river survival (V_n) and upstream passage survival. The null hypothesis (H0) represents no change from base period of 1975-1990 brood years. H1 and H2 represent hydro hypothesis and Stock Viability hypothesis for extra mortality, respectively. Plotted values are for illustration purposes.

Upper Columbia stocks could be incorporated as a third regional block to provide additional spatial and temporal contrast to Snake and Lower Columbia regions. One potential confounding factor is that the first step (A3/A5) restores free-flowing conditions in the lower Snake River *and* eliminates transportation

from McNary Dam. Assuming that McNary transportation is neutral to Upper Columbia spring chinook, hypothesized regional stock responses for H0 and H1 would be represented by Figure A.6-2.

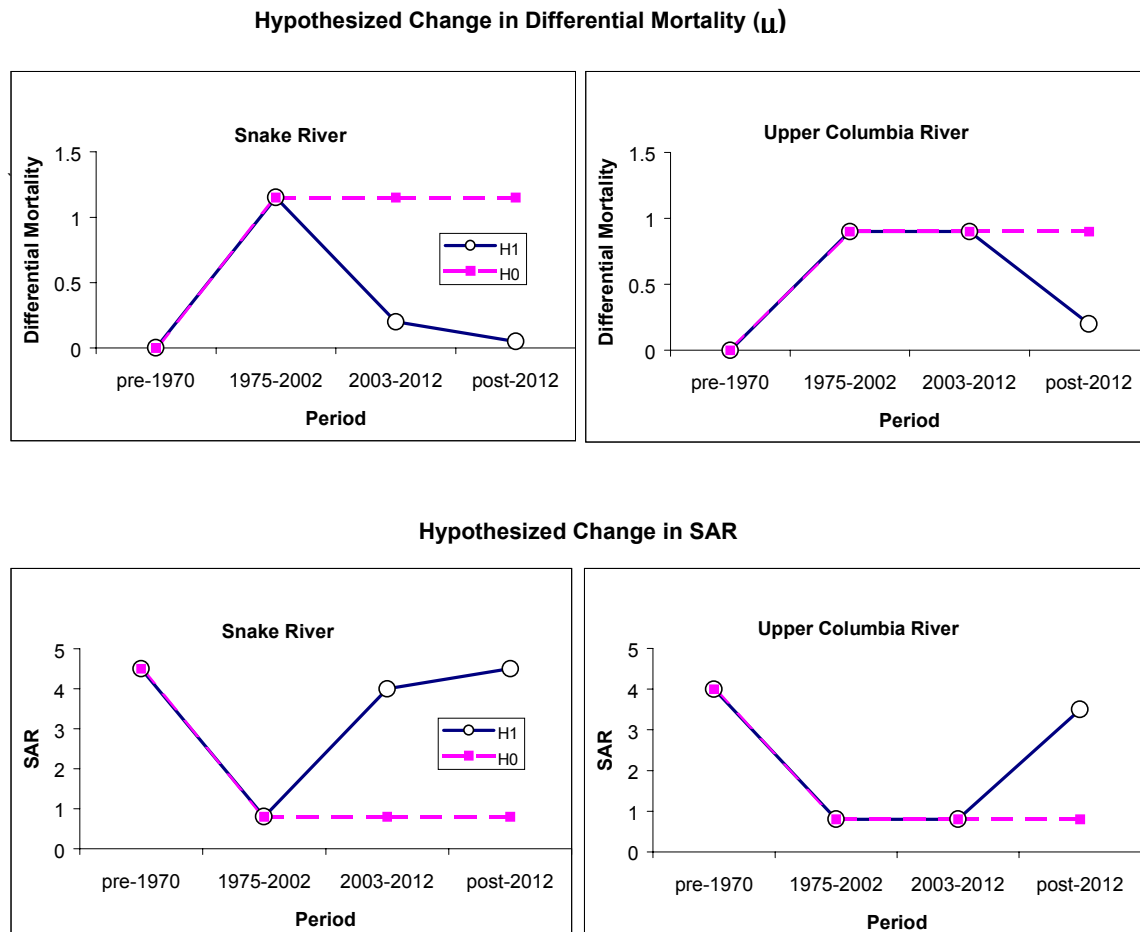


Figure A.6-2: Hypothesized changes in μ and SAR for stream-type chinook from the Snake River and Upper Columbia in response to implementing A3 in 2003 and B1 in 2012. The null hypothesis (H0) assumes no change from base period of 1975-1990 brood years. H1 is represented by the hydro hypothesis for extra mortality. Plotted values are for illustration purposes.

In this case, no change would be hypothesized under H1 in Upper Columbia μ or SAR for the 2003-2012 period due to implementation of A3/A5, and improvements would be expected to follow John Day drawdown (B1/B2) in 2012 (Fig. A.6-2).

If McNary transportation is not neutral (i.e., is either beneficial or detrimental) to Upper Columbia stocks, the H1 response in 2003-2012 would be higher or lower than represented in Fig. A.6-2. Two ways to deal with this potential confounding would be to explicitly hypothesize the effect of ceasing McNary transportation for Upper Columbia stocks, or to experimentally turn on and off transportation from this location.

A.6.2 Monitoring approach

Variables to monitor

Key variables to monitor are R/S (stream-type chinook and ocean-type chinook), and SAR (stream-type chinook and steelhead) for stocks in the three regions (Snake, Upper Columbia, Lower Columbia). R/S data require estimates of age-structured escapement, hatchery fractions on the spawning grounds, upstream passage loss, and harvest rates in the intercepting fisheries. SAR data require estimates of smolt numbers, age-structured adult returns, upstream passage loss, and harvest rates in the intercepting fisheries.

For stream-type chinook, we are interested in changes in the differential mortality between stock groups. From R/S data differential mortality has been expressed as μ (Deriso et al. 1996), and represents both direct and extra (or delayed) mortality. An analogous differential mortality statistic, $[-\ln(\text{SAR}_1/\text{SAR}_2)]$, could be developed for SAR data from the three regions (where subscripts represent different regions). Available SAR data from Warm Springs River and Yakima River (above 2 and 4 dams, respectively), indicate substantially better survival through this life stage for these stocks than for Snake River stocks.

For ocean-type chinook, we are primarily interested in differential changes in R/S patterns. (μ cannot be estimated because of lack of replication within region). SAR data may be difficult to obtain because of difficult logistics in sampling subyearlings at the same life stage (migration vs. rearing).

For steelhead, we are primarily interested in changes in differential SAR between regions. R/S data are scarce, due to more complex life-history patterns (e.g., variable smolt ages), and difficulty in accurately sampling spawning population sizes. Currently we have historic SAR estimates for aggregate wild runs from the Snake and upper Columbia.

To apply equations D-9 and D-16, an estimate of smolt survival [V_d] for downriver stocks (i.e., John Day) is also needed. To determine *where* net benefits in survival improvement may have occurred for Snake River stocks following dam breaching, would also require estimates of system survival and its components M (direct survival), D (differential survival of transported smolts post-BON), and P[b] for the period before breaching. Retrospective estimates of these parameters have been made in PATH using alternative passage models and hypotheses. Errors in estimating these quantities (particularly D) may make it very difficult to get accurate estimates of system survival.

Duration and intensity of monitoring

Frequency of sampling is annual. A long-term commitment should be made to collect R/S and SAR data throughout the Columbia Basin for this and other experimental management options.

Index stock R/S data need to be continued, and specific recommendations developed to improve future data collection (e.g., age composition, redd expansions, hatchery fraction accounting). A coordinated program would be developed to estimate SAR for steelhead and stream-type chinook index populations throughout the interior Columbia Basin.

A.6.3 Benefits, risks, costs, and trade-offs

Benefits and amount of learning possible

This approach implements of the least risky management action (natural-river restoration), within an experimental framework. The approach directly tests the outcome of implementing the best biological option for Snake River stocks, to apply results to decisions for Upper Columbia stocks.

The natural river options are the most likely to recover listed Snake River salmon, and are less risky than transportation options, according to PATH FY98 analyses. The natural river options exceeded all three standards used by NMFS to determine jeopardy for Snake River spring/summer and fall chinook salmon, with one exception. The likelihood of survival of spring/summer chinook missed the 24-year survival standard by less than one percentage point when breaching was delayed for eight years. In most cases, the natural river options met the standards under the most pessimistic assumptions. None of the transportation options met the recovery standard, except under very optimistic assumptions. NMFS' (1999) A-Fish sensitivity analysis (using PATH results and different assumed values of D) indicates that the natural river options outperform transportation, *except when high D-values are combined with non-hydro hypotheses about extra mortality*. (i.e., a high D-value combined with hydro-related delayed mortality of in-river fish still results in the best option being natural river).

Implementation of this action would aid decisions on whether to restore natural river conditions in the John Day pool reach for listed salmon and steelhead in the Upper Columbia River. The staggered decision points for Snake River drawdown and John Day drawdown lend themselves to a staircase design, if implementation follows the same temporal pattern. Delaying Snake River actions while studies are conducted on John Day would negate this time step.

This approach was previously described in a concept paper (Petrosky et al. 1998) submitted to the Multi-Species Framework Process in November 1998. In addition to benefiting listed anadromous stocks in the Snake and Upper Columbia, this approach would help restore ecosystem function and benefit native lamprey, white sturgeon, and resident fish and wildlife, and non-listed anadromous stocks from above John Day pool (ibid.).

Quantitative assessment of likely power to detect effects should be determined in FY2000. Because the desired effect size is large for total mortality reduction, and was estimable retrospectively, there is reason to believe that proposed monitoring could detect the desired effect. However, PATH has not investigated whether there would be sufficient power to clearly isolate which of the extra mortality hypotheses was more likely, given the future, observed regional stock responses.

Risks to stocks

SNAKE RIVER spring/summer chinook salmon are at extreme risk. Spawning population numbers since 1980 have been extremely depressed, and some spawning areas (Sulphur, Marsh creeks) have been devoid of spawners in some recent years. A greater concern is the fact that the depressed populations have been in decline since a brief positive trend during the early 1980s. For the seven stocks, the geometric mean of recruits per spawner to the spawning grounds (spawner to spawner ratio) has been less than 1.0 every year from 1984 through 1993 brood years (Fig. A.6-3). Since 1984, the geometric mean spawner/spawner ratio for the seven Snake River index stocks has been 0.44, that is, each generation has returned less than one-half the spawners of the previous generation. Obviously, populations cannot survive this trend indefinitely.

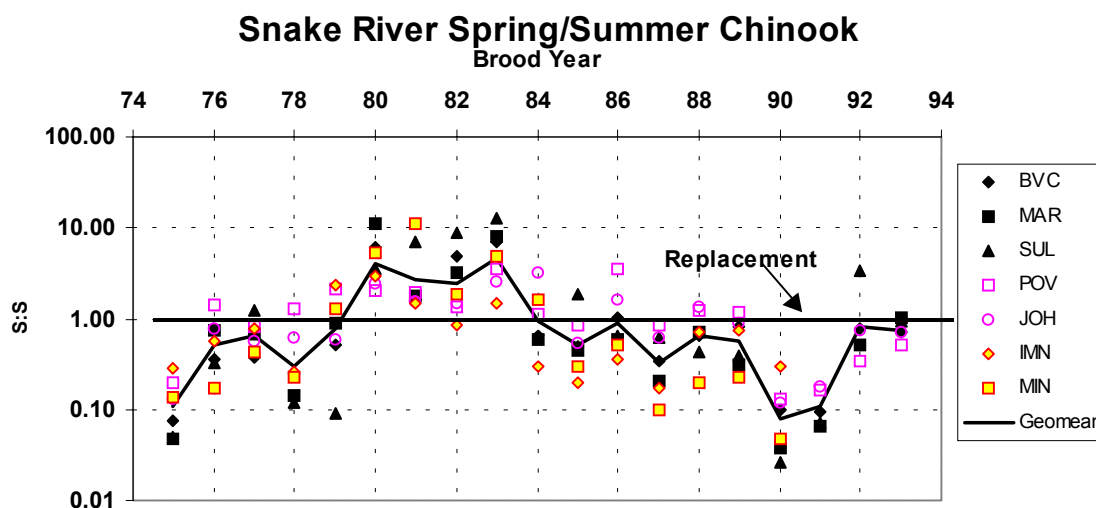


Figure A.6-3: Spawner to spawner ratios (S:S; natural log scale) for seven index stocks of Snake River spring/summer chinook, 1975-1993 brood years (Beamesderfer et al. 1997; PATH updates for brood years 1991-1993). A value less than 1.0 indicates that the population has declined for that brood year. Stocks are: Bear Valley (BVC), Marsh (MAR), Sulphur (SUL), Poverty Flat (POV), Johnson (JOH), Imnaha (IMN), Minam (MIN). S:S estimates not completed for Imnaha and Minam stocks for brood years 1991-1993.

Experimental management options that propose continuation of status quo hydropower operations, while studying components of extra mortality, need to explicitly recognize this risk. The proposed option prioritizes recovery to listed Snake River populations with the *least risky* hydrosystem alternative, and uses information gained to evaluate feasibility of natural river restoration through the John Day Pool reach specifically for upper Columbia stocks.

According to PATH FY98 the A3/B1 option has the lowest risk, and highest biological benefits of any of the experimental actions proposed. Transportation-based actions had lower probabilities of meeting survival and recovery standards, and were less robust to uncertainties. The decision analysis indicates that there is relatively less risk with the natural river options of falsely assuming the wrong extra mortality hypothesis for Snake River stocks. That is, recovery is likely for natural river options, regardless of which extra mortality hypothesis is correct.

Costs

Implementation costs for A3/B1 or A5/B2 options will be determined by the Drawdown Regional Economic Workgroup (DREW).

Costs of the proposed experimental management program have not been estimated. However, costs of a program to systematically evaluate responses in recruitment patterns and SARs to Snake River dam breaching would seemingly be comparable to current research, monitoring and evaluation efforts. Such a systematic program is needed to assess any of the long-term hydropower operations.

Trade-offs

Relative benefits are high and relative risks are low. Implementation costs for A3/B1 or A5/B2 options will be determined by DREW. Evaluation costs are expected to be similar to the current efforts.

A.6.4 Inferences

Table A-22: Example inference table for Hydro Extra Mortality Hypothesis.

Variable	Observation and Inference	
	Observations consistent with "Hydro"	Observations not consistent with "Hydro"
μ	Response consistent with that projected by H1 vs. H2 (Figure A.6-1 and A.6-2)	Response not consistent with that projected by H1 vs. H2 (Figure A.6-1 and A.6-2)
SAR or R/S	Response consistent with that projected by H1 vs. H2 (Figure A.6-1 and A.6-2)	Response consistent with that projected by H1 vs. H2 (Figure A.6-1 and A.6-2)
V_n	must be monitored but not directly relevant to testing extra mortality hypothesis	
λ_n	Increase	-*
$\Delta \lambda_n$	>1	$\leq 1^*$

* Critical observations; - ≈ no change

A.6.5 Confounding factors

In addition to the action of breaching dams, survival improvements potentially could be attributed to elimination of transportation, climate change, changes in passage survival at remaining dams, and/or changes in hatchery effects, etc.

The issue of confounding, and approaches to reduce it, will be examined in FY2000. In principle, it might be possible to pulse treatments for some of the potential confounding factors, such as hatchery production or transportation from McNary Dam. Confounding also might be reduced with explicit and quantitative, *a priori* statements of expected effects for Snake River spring/summer chinook, fall chinook and steelhead. For example, the hatchery hypothesis for extra mortality presumably does not apply to Snake River fall chinook, since they migrate after the hatchery spring migrants have departed. Potential changes in hatchery production that may tend to confound spring/summer response would not confound fall chinook response. Similarly, it does not seem likely that climate change would be selectively influential for both Snake River spring/summer chinook and fall chinook (compared to lower river stocks), since these stream-type and ocean-type stocks do not share in time and space the same estuary/ocean environments.

A.6.6 Practical Constraints

Implementation of natural river options would require congressional authorization, whether or not the actions are organized into an experimental management design. Assuming that natural river restoration actions would be authorized, there appear to be no serious logistical constraints to a program that systematically evaluates recruitment patterns and SARs from the Snake, upper Columbia and lower Columbia regions. Costs of such a program would seemingly be comparable to current research, monitoring and evaluation efforts.

The initial decision on whether to pursue natural river options for listed Snake River stocks will come with NMFS's biological opinion on the operation of the Columbia Basin hydroelectric system in 1999. A review of the biological, economic and legal case for natural river options (Blumm et al. 1999) concludes that breaching of Snake River dams is economically affordable based on several economic studies, and that this option would produce net social benefits.

While achieving congressional authorization may be difficult, Blumm et al. argue that continuation of the status quo FCRPS operations is "legally unacceptable" on several grounds, and that legal processes in addition to ESA may come into play:

"Although ESA will dominate the legal landscape during the next couple of years, the Northwest Power Act, the Federal Power Act, the Clean Water Act, Indian treaty fishing rights and the Pacific Salmon Treaty could also affect the drawdown decision." (p. 132).

"Among the largest legal threats to the current status quo in Idaho is the potential demand for water to restore Snake River salmon runs, either to satisfy the ESA, the Clean Water Act, or the Nez Perce Tribe's reserved water rights to the Snake River. Because these claims are quite large, they could jeopardize the water rights of numerous upstream diverters... Settling these claims through enactment of federal legislation authorizing breaching of lower Snake River dams and lowering John Day reservoir offers the best chance of restoring the fishing economy of both the Nez Perce Tribe and the state of Idaho, while preserving irrigation economies of Idaho and eastern Oregon and Washington." (p. 153).

Literature Cited

Blumm, M.C., L.J. Lucas, D.B. Miller, D.J. Rohlf and G.H. Spain. 1999. Saving Snake River water and salmon simultaneously: the biological, economic and legal case for breaching the lower Snake River dams, lowering John Day Reservoir, and restoring natural river flows. *Environmental Law* 28(4):101-153.

Petrosky, C., H. Schaller, P. Wilson, E. Weber and O. Langness. 1998. Integration of ESA recovery actions and experimental management into a multi-species framework. Multi-Species Framework Concept Paper, November 17, 1998 Workshop. Northwest Power Planning Council. Portland, Oregon.

Appendix B. Bayesian Approach to Evaluating Learning

This appendix describes a formal Bayesian updating procedure, where probabilities on alternative hypotheses change over time as data is collected. Actions that do not provide good learning opportunities (i.e., their effects are not detectable) will result in smaller changes in the posterior probabilities on the hypotheses than actions where effects are detectable.

Formally, the probability placed on hypothesis “a” in any time period during the experimental period given that hypothesis j is true is:

$$\Pr_{t+1}(H_a|H_j) = \frac{L(\text{data}_{t+1}|H_a) * P_t(H_a)}{\sum_i L(\text{data}_{t+1}|H_i) * P_t(H_i)} \quad (3)$$

One can then calculate the probability of recovery (given that one of the hypotheses is true) associated with an experimental action as new data are collected.

$$\Pr_{t+1}(\text{Recovery}|H_j) = \sum_i \Pr(\text{Recovery}_{t+1}|H_i) * P_{t+1}(H_i) \quad (4)$$

The probability of recovery for an action over all possible hypotheses is

$$\Pr_{t+1}(\text{Recovery}|Action\ k) = \sum_i \Pr_{t+1}(\text{Recovery}|H_j) * P_t(H_j) \quad (5)$$

Appendix C. Reevaluation of the method used to predict SARs from Recruits/Spawner

Smolt to adult survival rates (SARs) are computed based on an implied relationship between historical R/S and SAR estimates:

$$\ln(\text{SAR}) = q + a_i + m_t + \varepsilon'_t$$

where q is a proportionality constant estimated from SAR and R/S data.

This approach is analogous to how the BSM presently calculates SARs. When the methods behind this approach were derived (FY97 Analysis report, May, 1998, Section 3, Chapter 9) it was assumed that the Raymond SAR estimates (passage years 1962-84) were largely independent of the S/R data for the index stocks. However, it turns out that adults estimated by Raymond (the "A" term in SAR) are strongly correlated with recruits for the index stocks (Pearson "r" of 0.9 to 0.95). In part, this is because the R/S "R" makes up about 10-30% of the Raymond "A". [AS a 'thought experiment,' imagine the case where the index stocks comprised 100% of the spring/summer chinook over Lower Granite]. In any case, the assumption that the SAR and R/S are independent of each other seems very tenuous. In addition, Raymond-style SARs are roughly an order of magnitude higher than concurrent SARs calculated from coded wire tags or PIT tags (C. Toole, phone conversation of 11/24/99). Since in future SARs would presumably be estimated using returns of PIT-tagged fish, this casts some doubt on methods which rely on forecasts of Raymond-style SAR measurements, which depend on an estimate of the total number of smolts passing Lower Granite and the returns to LGR of adults derived from those smolts.

Assuming that the above-noted problems are indeed correct, this in turn means that we cannot say what the strength of the relationship may be between PIT-tag based SARs and R/S. Therefore, if we try to evaluate the R/S resulting from a management action that should increase SARs (e.g., modify transport and measure "D"), we have little or no data in hand to say what the relationship between PIT-tag derived SARs and R/S might be. All we can do for now is to assume various degrees of correlation and run the power analyses accordingly. Perhaps a task for the next few weeks might be a systematic comparison of Raymond-style, T/C experiment, and PIT-tag based SAR estimates, to assay the relationships among them.

Appendix D. A graphic contrast of hatchery steelhead abundance and spring chinook SARs for 1990 through 1995

Introduction

I plotted the weekly average passage indices for spring/summer chinook (combined hatchery and wild) and hatchery steelhead along with the combined hatchery and wild SARs for spring chinook to determine if there was a visual relationship between the abundance of hatchery steelhead and the survival of spring/summer chinook. The purpose was to explore the feasibility of reducing the production hatchery steelhead as a means of increasing spring/summer chinook survival to levels consistent with ESA survival and recovery goals.

Methods

Passage indices (PIs) for spring/summer chinook (hatchery and wild combined) and hatchery steelhead were provided by Penelope Sanders, Fish Passage Center, for 1990 through 1995. Alan Byrne, Idaho Fish and Game, provided weekly SAR data for the same years for both wild and hatchery chinook tagged both above and at Lower Granite Dam. I pooled all the data to maximize weekly sample sizes. The 1990 – 1995 period was chosen because prior to 1990 hatchery and wild steelhead were not distinguished in samples. In 1996 data plots showed that PIs did not provide the desired contrast early in the year when spring chinook are becoming increasingly abundant while hatchery steelhead are either not present or at least not abundant. 1997 and 1998 are more promising in that regard but cohorts are incomplete.

Note that in two years, 1993 and 1994, PIs began on April 15; SARs of zero are due to a lack of fish, not poor survival, in those years. Also, the 90% passage completion date for chinook typically falls in late May and in some years sample sizes beyond that point become too small to be useful.

Results

In contrasting annual plots of weekly average PIs and SARs for 1990 through 1995 (see Figures D-1 to D-6) no clear relationship between the abundance of hatchery steelhead and spring/summer chinook emerges. Low SARs at the onset of spring/summer chinook annual migrations are probably not due to hatchery steelhead whose migrations typically don't begin for two or three weeks. In some years low abundance of steelhead resulted in slight elevations in chinook SARs early in the migration season but in other years SARs were extremely low despite an apparent near absence of hatchery steelhead. In 1995, the year with by far the most tagged fish, chinook SARs increased as the abundance of hatchery steelhead rose and did not decline until the steelhead abundance dropped.

In half the years (1992, 1993 and 1994) a modest rise in chinook SARs was followed by decreases later on when steelhead abundance increased. But in other half (1990, 1991 and 1995) the chinook SARs increased as hatchery steelhead increased in abundance.

Also, under no conditions in any year did the SARs approach the two percent minimum goal established by PATH over the course of the season, regardless of steelhead abundance.

Discussion

Visual observations provide no relationship between spring/summer chinook survival and hatchery steelhead abundance. There are periods of low steelhead abundance with extremely low chinook SARs as well as periods of relatively high steelhead abundance accompanied by relatively high chinook SARs.

Regardless of steelhead abundance, transported spring/summer chinook survived at low rates. Only in three weeks within the six years did the SARs meet or exceed one percent. Note that these SARs are from Lower Granite Dam only where transport survival is typically the highest. Lower dams such as Lower Monumental and McNary, if added to this type of analysis, would show even lower SARs for chinook even though steelhead would presumably be present in lower relative, and absolute abundance, because of higher FGEs for steelhead.

While it is certainly realistic to suspect that hatchery steelhead could consume, injure or at least stress the smaller spring/summer chinook, there is nothing in the visual observations to indicate that even the total elimination of the steelhead hatchery program would result in the restoration of spring/summer chinook to the level of survival (approximately two percent), far less recovery (approximately four percent.) Thus while hatchery steelhead likely contribute to the poor performance of spring/summer chinook, there are clearly other contributors and these other contributors would appear to pose greater limitations to chinook recovery than hatchery steelhead.

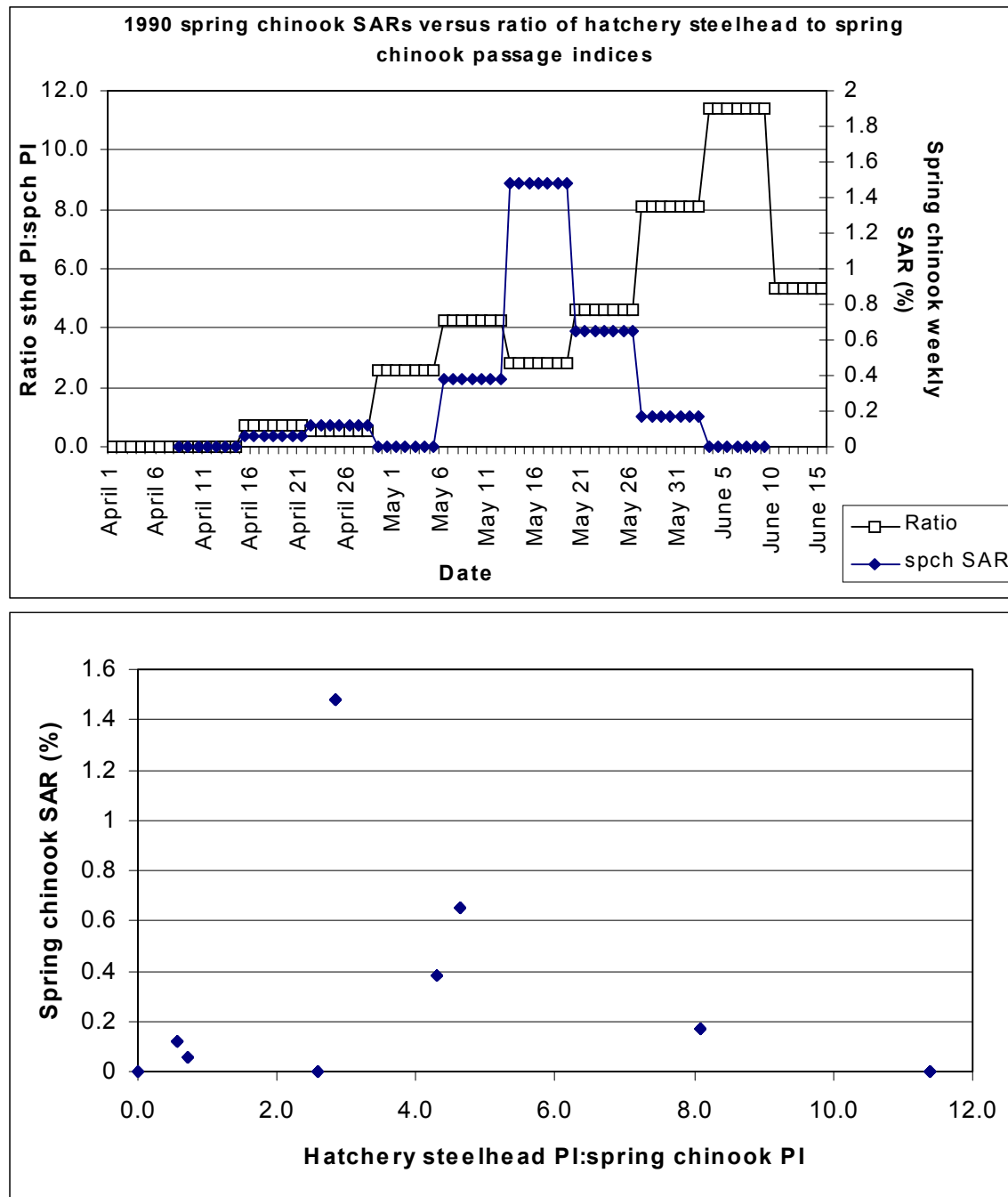


Figure D-1. 1990 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

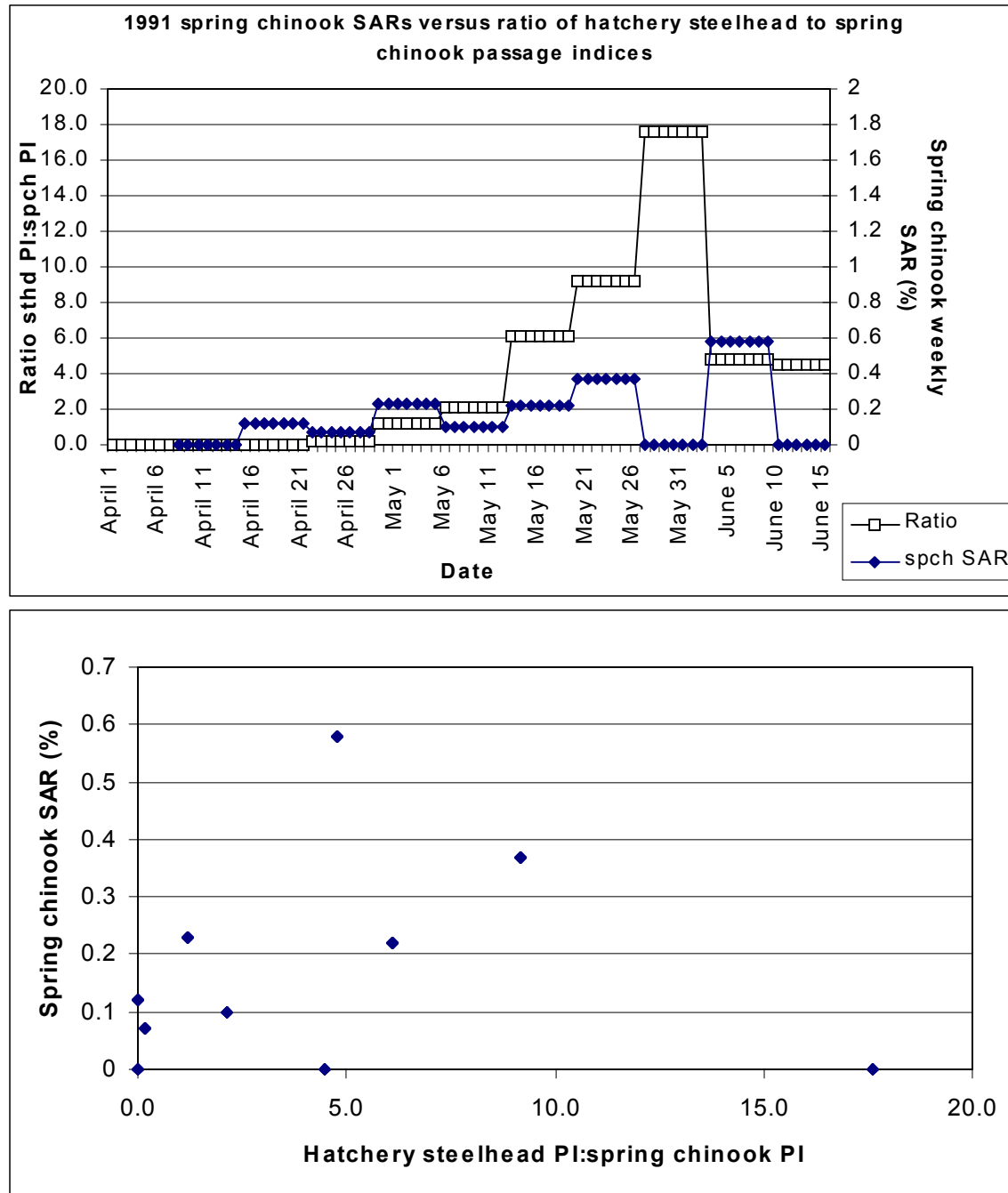


Figure D-2. 1991 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

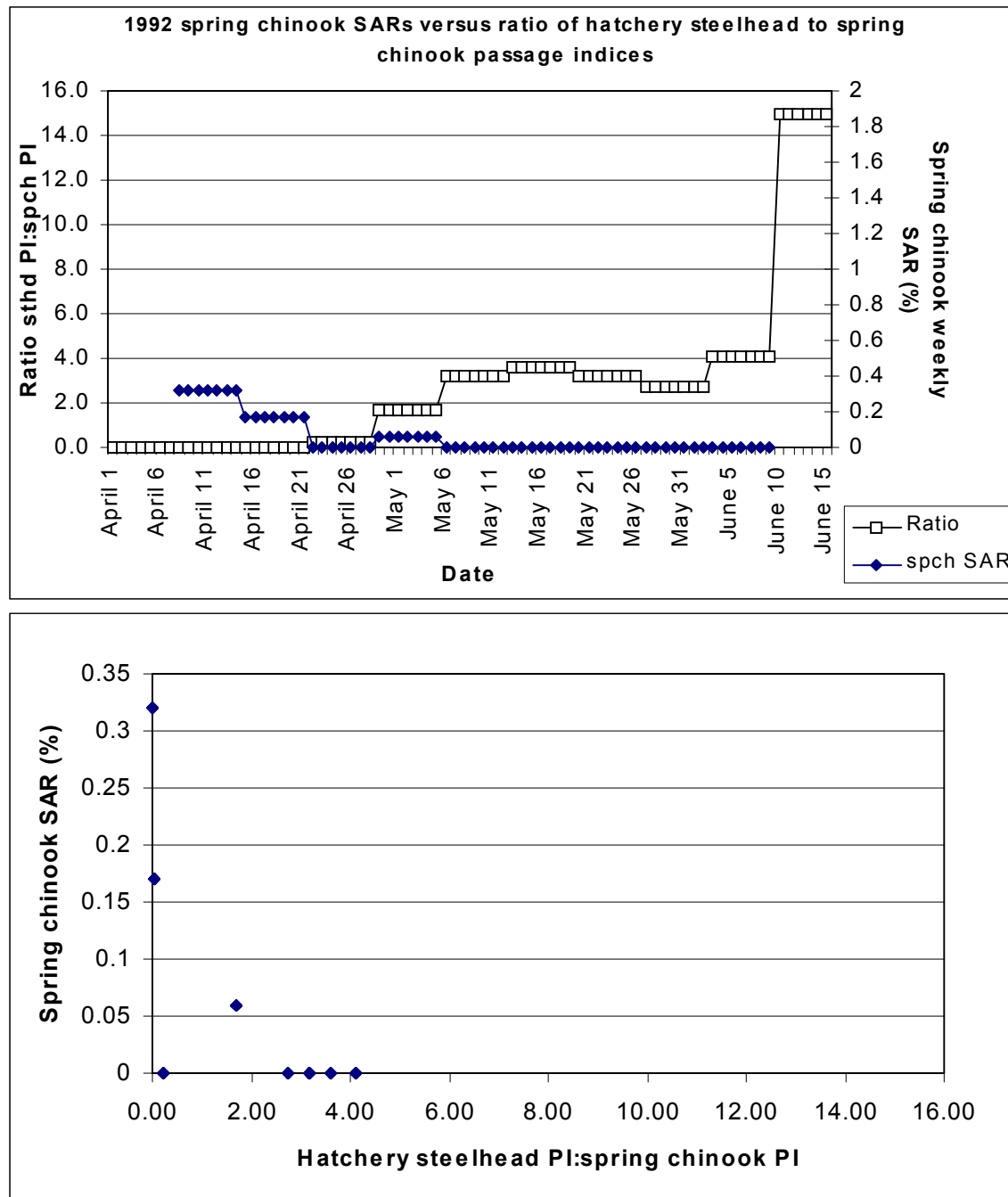


Figure D-3. 1992 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

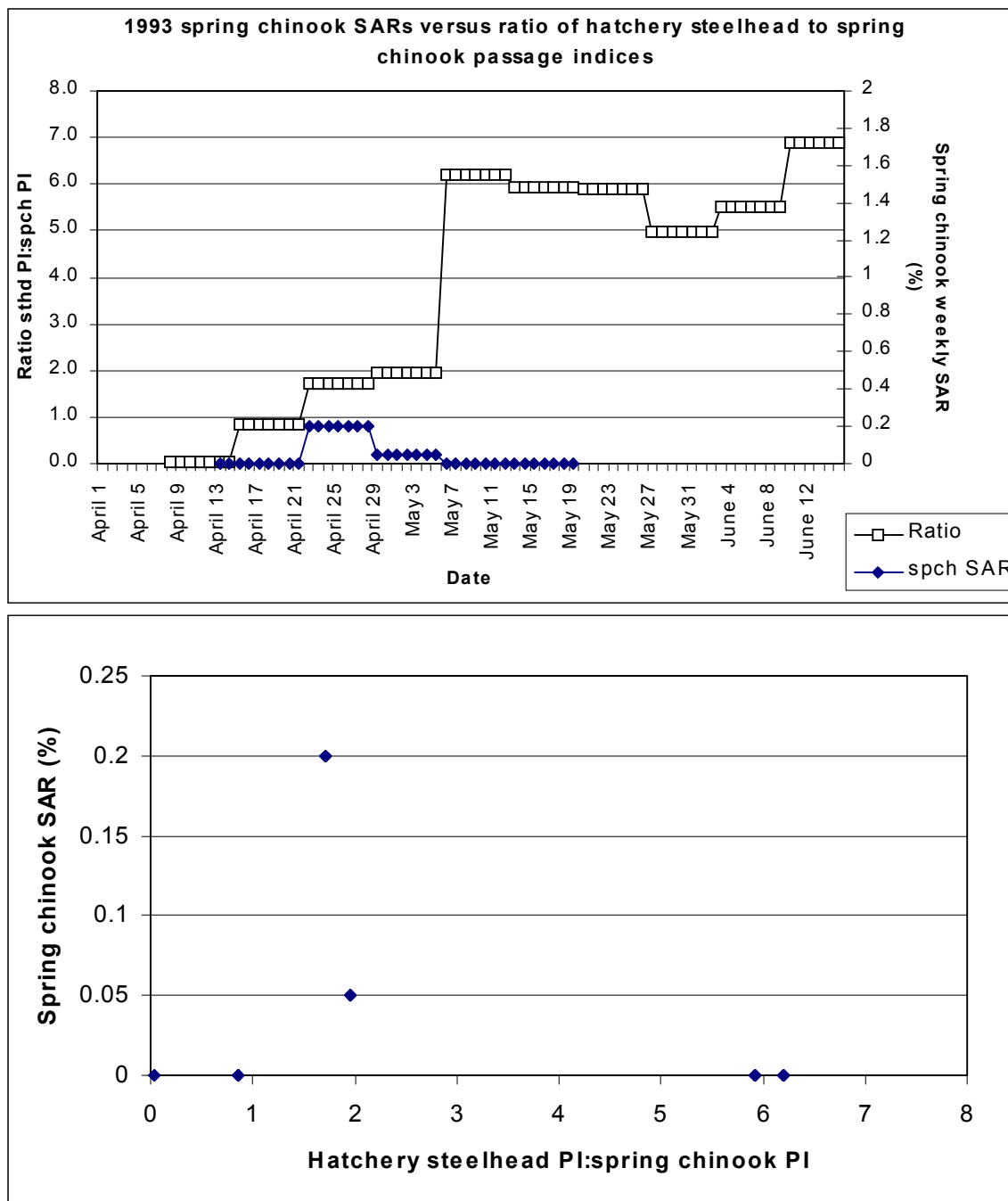


Figure D-4. 1993 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

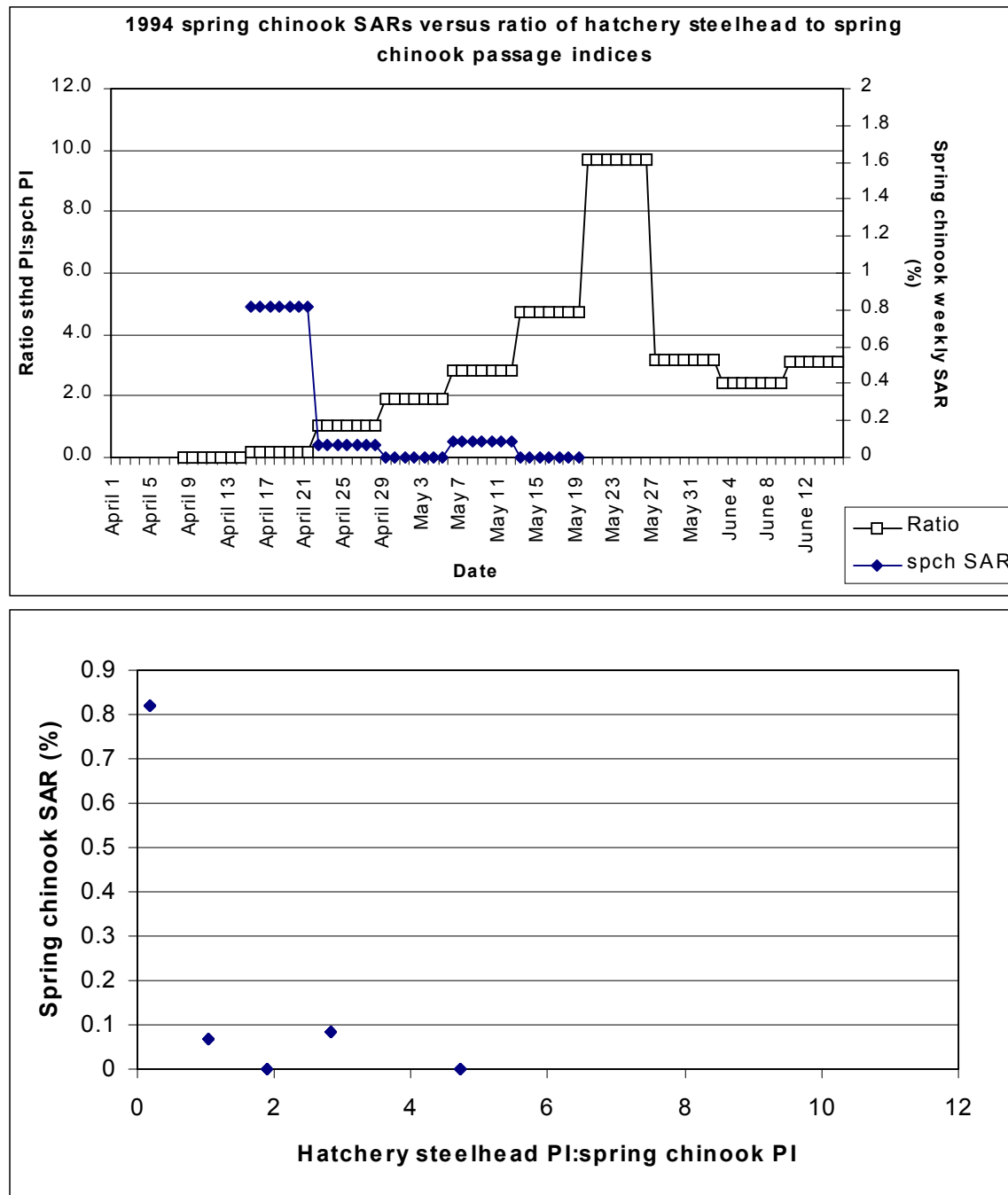


Figure D-5. 1994 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

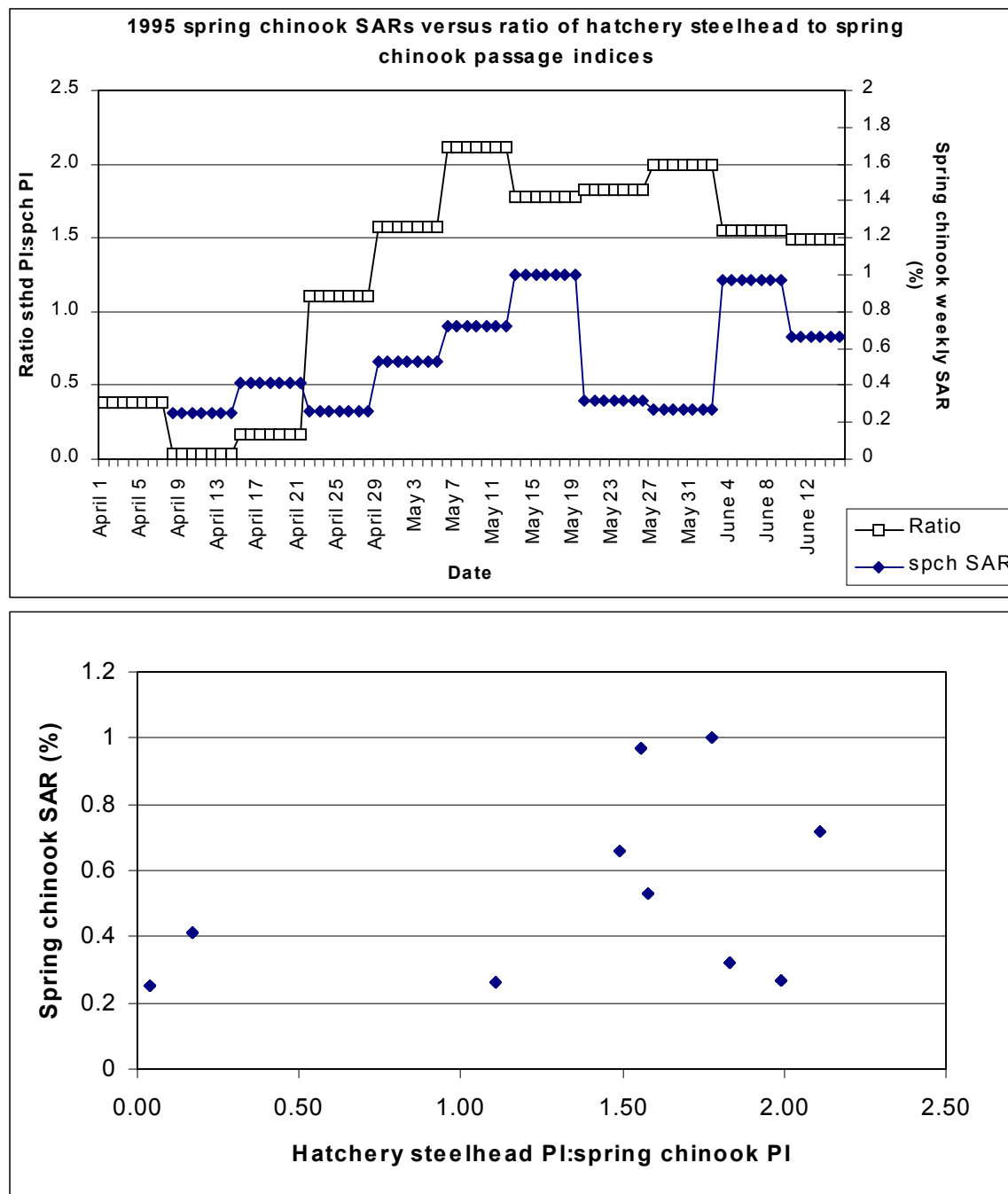


Figure D-6. 1995 Spring/summer chinook and hatchery steelhead passage indices and chinook SARs.

Appendix E. Population Projections

In this appendix we present the jeopardy standard and quasi-extinction metrics for all actions and all stocks. In order to calculate these metrics, we assume that all actions will be maintained for the duration of each metric's time horizon (i.e. 24 and 100 years for survival probabilities, 24 and 48 years for recovery probabilities, and 10 and 100 years for quasi-extinction metrics). With the possible exception of the drawdown actions, this assumption is probably not realistic because if one discovers a suite of actions that works in the sense of meeting survival and recovery requirements (however these are determined), one likely would not continue with the original on/off experiment. Instead, one would either decide on a "final" course of action or modify the action(s) and monitoring scheme(s) based on newly acquired information. However, a formal analysis of this type of multi-stage decision analysis is beyond the scope of the current report. The population metrics included here may thus be viewed as a relative index of the biological effects of the actions on the stocks, if the actions were continued indefinitely.

Population Measure	Stock	Base Case	Generic Actions						Experimental Actions													
			#1	#2	#3	#4	#5	#6	Modify Trans.	Trans on/off	Carcass #1 $\Delta m=0.2$ treatmnts constant	Carcass #2 $\Delta m=0.7$ treatmnts constant	Carcass #3 $\Delta m=0.7$ treatmnts vary	Hatchery Reduction	Drawdown D=0.3/0.5, 3-year		Drawdown D=0.3/0.5, 8-year		Drawdown D=0.8/0.8, 3-Year		Drawdown D=0.8/0.8, 8-Year	
			0/1 on/off	#1 w/ uniform dist.	#1 w/ meas. err off	0/1 5-on, 5-off	#1 w/ Delta Model	0/1 for 10 yrs, then 1							BKD	Hydro	BKD	Hydro	BKD	Hydro	BKD	Hydro
24-Year Survival	Imnaha	0.408	0.65	0.66	0.66	0.60	0.74	0.70	0.51	0.28	0.51	0.69	0.59	0.71	0.69	0.73	0.58	0.61	0.53	0.73	0.48	0.61
	Minam	0.479	0.67	0.68	0.68	0.63	0.64	0.72	0.55	0.34	0.47	0.44	0.61	0.72	0.71	0.74	0.62	0.65	0.58	0.74	0.54	0.65
	Bear	0.285	0.52	0.51	0.51	0.47	0.62	0.57	0.37	0.18	0.38	0.57	0.45	0.58	0.57	0.62	0.45	0.49	0.40	0.62	0.35	0.49
	Marsh	0.151	0.33	0.33	0.33	0.28	0.54	0.39	0.21	0.09	0.15	0.13	0.25	0.40	0.39	0.47	0.28	0.32	0.23	0.47	0.19	0.32
	Sulphur	0.172	0.35	0.35	0.35	0.31	0.49	0.41	0.23	0.10	0.24	0.40	0.29	0.41	0.41	0.47	0.31	0.35	0.26	0.47	0.22	0.35
	Poverty	0.297	0.54	0.55	0.54	0.48	0.67	0.61	0.37	0.19	0.29	0.29	0.45	0.62	0.60	0.66	0.46	0.51	0.40	0.66	0.35	0.51
	Johnson	0.283	0.52	0.52	0.52	0.47	0.58	0.58	0.37	0.18	0.37	0.57	0.45	0.58	0.58	0.63	0.46	0.50	0.40	0.63	0.35	0.50
100-Year Survival	Imnaha	0.377	0.76	0.75	0.77	0.70	0.84	0.87	0.54	0.16	0.55	0.82	0.67	0.82	0.88	0.91	0.85	0.88	0.66	0.91	0.65	0.88
	Minam	0.455	0.74	0.74	0.75	0.70	0.77	0.84	0.58	0.23	0.43	0.36	0.67	0.80	0.86	0.88	0.83	0.86	0.67	0.88	0.66	0.86
	Bear	0.336	0.71	0.70	0.72	0.67	0.82	0.83	0.50	0.12	0.51	0.79	0.62	0.79	0.85	0.89	0.82	0.85	0.62	0.89	0.61	0.85
	Marsh	0.161	0.55	0.56	0.56	0.50	0.76	0.74	0.30	0.04	0.14	0.09	0.42	0.68	0.78	0.83	0.74	0.79	0.43	0.83	0.41	0.79
	Sulphur	0.224	0.51	0.50	0.52	0.48	0.64	0.66	0.34	0.08	0.35	0.60	0.43	0.59	0.69	0.74	0.66	0.71	0.44	0.74	0.42	0.71
	Poverty	0.251	0.67	0.67	0.68	0.61	0.77	0.82	0.41	0.08	0.24	0.23	0.55	0.76	0.85	0.89	0.81	0.85	0.55	0.89	0.53	0.85
	Johnson	0.274	0.63	0.62	0.64	0.57	0.69	0.76	0.42	0.10	0.42	0.69	0.53	0.70	0.79	0.83	0.76	0.80	0.53	0.83	0.52	0.80
24-Year Recovery	Imnaha	0.019	0.28	0.30	0.29	0.24	0.56	0.63	0.06	0.00	0.07	0.43	0.16	0.46	0.77	0.92	0.58	0.80	0.14	0.92	0.10	0.80
	Minam	0.022	0.27	0.26	0.28	0.24	0.37	0.56	0.06	0.00	0.01	0.00	0.14	0.40	0.70	0.86	0.57	0.77	0.15	0.86	0.11	0.77
	Bear	0.008	0.18	0.19	0.16	0.14	0.41	0.44	0.04	0.00	0.04	0.33	0.09	0.33	0.60	0.87	0.36	0.58	0.07	0.87	0.05	0.58
	Marsh	0.001	0.05	0.05	0.04	0.03	0.29	0.17	0.01	0.00	0.00	0.00	0.02	0.11	0.26	0.54	0.09	0.22	0.01	0.54	0.01	0.22
	Sulphur	0.005	0.11	0.12	0.10	0.09	0.30	0.30	0.02	0.00	0.02	0.20	0.05	0.20	0.40	0.64	0.25	0.44	0.04	0.64	0.03	0.44
	Poverty	0.006	0.17	0.18	0.16	0.12	0.38	0.46	0.03	0.00	0.01	0.01	0.07	0.31	0.65	0.90	0.38	0.64	0.06	0.90	0.04	0.64
	Johnson	0.019	0.29	0.30	0.29	0.24	0.43	0.63	0.05	0.00	0.06	0.43	0.16	0.43	0.77	0.90	0.59	0.80	0.13	0.90	0.10	0.80
48-Year Recovery	Imnaha	0.018	0.32	0.30	0.32	0.26	0.59	0.70	0.06	0.00	0.07	0.48	0.17	0.49	0.81	0.95	0.82	0.95	0.17	0.95	0.17	0.95
	Minam	0.021	0.25	0.24	0.25	0.20	0.43	0.58	0.08	0.00	0.02	0.00	0.15	0.36	0.71	0.84	0.71	0.84	0.15	0.84	0.15	0.84
	Bear	0.012	0.23	0.22	0.23	0.21	0.57	0.64	0.05	0.00	0.05	0.45	0.12	0.41	0.78	0.92	0.78	0.93	0.14	0.92	0.13	0.93
	Marsh	0.003	0.14	0.12	0.13	0.12	0.45	0.53	0.02	0.00	0.00	0.00	0.06	0.30	0.68	0.87	0.66	0.86	0.07	0.87	0.06	0.86
	Sulphur	0.008	0.15	0.15	0.15	0.14	0.36	0.45	0.03	0.00	0.04	0.30	0.08	0.26	0.56	0.72	0.56	0.72	0.09	0.72	0.09	0.72
	Poverty	0.005	0.20	0.21	0.20	0.16	0.44	0.61	0.02	0.00	0.00	0.00	0.10	0.38	0.77	0.94	0.77	0.94	0.10	0.94	0.09	0.94
	Johnson	0.019	0.29	0.33	0.30	0.24	0.48	0.67	0.05	0.00	0.05	0.45	0.13	0.45	0.79	0.89	0.79	0.89	0.16	0.89	0.16	0.89
CRI 10-Year Quasi-Extinction	Imnaha	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Minam	0.001	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Bear	0.002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Marsh	0.655	0.66	0.62	0.66	0.67	0.16	0.66	0.67	0.69	0.67	0.67	0.66	0.65	0.67	0.66	0.67	0.67	0.67	0.66	0.67	0.67
	Sulphur	0.444	0.44	0.41	0.43	0.46	0.50	0.44	0.46	0.49	0.42	0.43	0.44	0.43	0.46	0.45	0.47	0.47	0.46	0.45	0.47	0.47
	Poverty	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Johnson	0.003	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01
CRI 100-Year Extinction	Imnaha	0.058	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.38	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Minam	0.072	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.45	0.08	0.13	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00
	Bear	0.083	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.51	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Marsh	0.874	0.73	0.69	0.72	0.74	0.25	0.71	0.80	0.98	0.89	0.92	0.75	0.71	0.71	0.71	0.73	0.73	0.76	0.71	0.77	0.73
	Sulphur	0.807	0.65	0.67	0.64	0.68	0.68	0.62	0.73	0.95	0.71	0.63	0.69	0.64	0.63	0.65	0.65	0.67	0.69	0.65	0.71	0.67
	Poverty	0.164	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.72	0.17	0.18	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
	Johnson	0.156	0.01	0.02	0.01	0.02	0.03	0.02	0.03	0.65	0.03	0.01	0.02	0.04	0.03	0.06	0.03	0.07	0.02	0.06	0.02	0.07

Appendix F. Exploration of hypothesis tests of true and realized D values

Sections A.1 and A.2 derive the number of PIT-tagged smolts needed to test particular hypotheses about ‘ D ’ with a desired amount of power. However, this analysis is designed for a one-project experiment, whereas Snake River chinook would be transported at 3 or 4 projects (A1 and A2, respectively), and assumes that only sampling error affects the ability to estimate a relevant D . In currently available PIT-tag data comparing SARs of transported and non-transported smolts, a significant number of fish were transported at four projects (LGR, LGS, LMN) in 1994 and three projects (LGR, LGS, LMN) in 1995 and 1996. The data suggest both that D differs depending on the project from which smolts are transported, and that inter-annual variance in D due to a combination of sampling and process error at a given project may be quite large (e.g. see Bouwes et al. 1999). An effect similar to the sampling effect of the subset of the aggregate Snake River population (the experimental sample) due to the small number of PIT-tagged adults returning to Lower Granite Dam, also applies at present to actual entire individual spawning stocks, since spawning escapements have been extremely low in recent years.

Variance in annual D ’s reflects variance in survival rates of a large component of the migrating population and so will affect the long term population growth rate of the index stocks over time (e.g. see ISAB 1999). A description of D that truly describes how transport affects the long-term fate of the stocks under A1 and A2 must take into account both inter-project and inter-annual variance in D values. The power to estimate the D values experienced by the transported population as a whole will then depend upon the relative proportions of the transported population which are collected and barged from each collector project, and the mean and variance in D from each collector project during the period of interest. The mean and variance of system survival rate (ω) will depend on these proportions and D values, as well as the overall survival rate of each category describing the separate groups of fish migrating through the hydrosystem.

Previously within PATH, system survival rate and D have been represented simply; only one transport group was kept track of for purposes of providing output to the life-cycle models. A more complete and accurate formulation of the hydrosystem experience can be obtained by explicitly expressing the experience of each group of smolts, the groupings being determined by their separate ‘destinies’. For example, with collection and transportation at three Snake River projects occurring, four different groups of smolts need to be tracked from their earliest encounter with the hydrosystem (i.e. the head of Lower Granite reservoir):

- the group destined not be transported (non-transported or in-river group)
- the group destined to be transported from LGR dam, the group destined to be transported from LGS dam
- the group destined to be transported from LMN dam
- the group destined to be transported from MCN dam, for scenarios that include transportation at MCN [e.g. A2]

The fate of the in-river group in a particular year can be described by one parameter, the in-river hydrosystem survival rate (V_n). The expressions for the fate of other groups are more complicated, and we must also express the portion of the migrating population at the start of the hydrosystem that is destined to pass by each of the routes in a year. Let’s call this portion parameter P_{0j} . The j ’s can take the value 1 to 5, for the first five hydroelectric projects encountered during the downstream migration, where (except for IHR) transportation is an option, or ‘ n ’ (for in-river fish). Then the proportion destined not to be transported in can be expressed as

$$P_{0,n} = 1 - \sum_{j=1}^5 P_{0,j}$$

Each destined-to-be-transported group will be subject to a certain mortality in-river until reaching its destined collector project. Presumably, the survival rate for each group migrating in-river over identical reaches will be very similar (this may not be case, but it simplifies the analysis and is more likely to be true than other simplifying assumptions). Each transport group j has a survival rate from the head of LGR pool until it reaches its unique point of collection (call it V_j), a survival rate on barge or truck from placement in the conveyance until release below Bonneville (which may differ between collector sites, as well as between years—call it $s_{b,j}$), and a ratio (relative to non-transported smolts) of survival rates from BON back to spawning (D value, more specifically D_j). Therefore, the ‘system survival’ of smolts can be expressed as

$$P_{0,n} * V_n + \sum_{j=1}^5 (P_{0,j} * V_j * s_{b,j} * D_j)$$

with $P_{0,j} = 0$ if transport isn’t occurring from project j .

We can use estimates of D from recent PIT tag data (Bouwes et al. 1999) together with passage model estimates of the other parameters to estimate more accurately the mean and variance of system survival and overall D value, for a given set of assumptions. If P_j is the fraction of smolts arriving at dam j that are transported, then the proportion destined to pass by route j can be calculated within the passage models according to

$$\begin{aligned} P_{0,j} &= P_j & \text{if } j = 1 \\ P_{0,j} &= P_j \cdot \prod_{k=1}^{j-1} (1 - P_k) & \text{if } j > 1 \end{aligned}$$

The $P_{0,j}$ can also be calculated from passage model $P_{b,j}$ values, where $P_{b,j}$ is the fraction of all of the smolts surviving to immediately below Bonneville dam that were transported from dam j . In this case,

$$P_{0,j} = \frac{\exp(-M) \cdot P_{b,j}}{s_{b,j} \cdot V_j}$$

where M is direct, instantaneous passage mortality.

Data for estimating mean and variance of D ’s used in this analysis come from the Bouwes et al. (1999) analysis of project specific D ’s for three years (Table F-1). The theoretical distribution of T/C’s (and hence D ’s, if variances in estimating control survival rates are ignored) is lognormal (Harmon et al. 1993). The geometric mean estimates the central tendency of a variable which fluctuates widely better than the arithmetic mean, and is appropriate for ratio scale data (Zar 1984).

Table F-1. *D* estimates from Bouwes et al. (1999) spreadsheet, wild spring/summer chinook, per-mile expansion

Project	1994	1995	1996
LGR	1.331	0.515	0.435
LGS	1.155	0.301	0.976
LMN	0.397	0.000	0.000
MCN	0.000	0.000	0.000

Geometric means for each project could be computed in a number of ways. The natural logarithms of the annual estimates could simply be averaged and the anti-logarithm of the average taken. However, for this exercise, I wished to weight each year's *D* estimate for a project by a measure of the precision of the estimate. The usual equation for variance of *T/C* or *D* (presented in Section A.1.2) would not work with these data, since the lack of adult returns from the transport group in some years results in an undefined variance. Variance of MCN *D* could not be estimated at all, since in the one year where many PIT-tagged fish were transported from MCN, none returned as adults. Very few were transported from MCN in 1995 and 1996 and none of these returned.

I chose instead to weight the mean and variance for the upper three project *D*'s using the harmonic mean of the number of smolts transported from a particular project and the number of tagged smolts released from LGR dam which migrated in a "run-of-the-river fashion" (i.e. were not detected at any of the collector projects) (Table F-2). This allowed circumventing the problem of lack of adult returns from the transport group in some years without omitting the information (of the upper three projects, this applied only at LMN). In no years were there zero non-transported returns. The method of using harmonic mean of releases also is more relevant to one of the primary tasks at hand: estimating inter-annual variance in *D*'s. Using the standard formula for variance to derive weights between year would attribute all variation in adult returns to *sampling* variance, so that years with fewer adult returns from transport (or non-transported) groups automatically get lower weight. In this case, the magnitude of inter-annual variance in *D* will be underestimated, if varying environmental conditions affect expected SAR values of either transported or non-transported smolts, or both. Finally, it should be noted that with such a small data set, the weighted mean and variance of *D* at each project with non-zero *D* estimates is potentially sensitive to the method used to weight each year's contribution.

Table F-2. Weighted averages, geometric means, and standard deviations by project (years weighted by harmonic mean of release numbers of transport and in-river groups).

Project	Weighted Average <i>D</i>	Weighted average $\ln(D)$	Weighted Geomean <i>D</i>	Weighted Std Dev of $\ln(D)$	Weighted Std Dev of <i>D</i>
LGR	0.795	-0.349	0.705	0.479	0.402
LGS	0.775	-0.438	0.645	0.648	0.406
LMN	0.247	NA	NA	NA	0.193
MCN	0	NA	NA	NA	NA

For the upper two projects, the geometric mean is significantly less than the arithmetic mean. The arithmetic mean and standard deviation of non-transformed *D* values are useful in creating lognormal distributions from which to draw randomly, in order to simulate a time series of variable *D*'s for each project. The method described in Burgman et al. (1993) can be used:

Estimate the mean (μ) and standard deviation (σ) of the non-transformed *D* estimates. Let $c = \sigma / \mu$, and compute $m = \ln(\mu) - \ln(c^2 + 1)$ and $s = \sqrt{\ln(c^2 + 1)}$. Sample a random number, *Y*, from the

normal distribution with mean m and standard deviation s . Then, the lognormal random number is $L = \exp(Y)$. Table F-3 shows the parameters used to create a lognormal distribution from which to create D 's using the above method.

Table F-3. Parameters for random draws of D by project, and expected value of D for each project from draw.

Project	Variable	μ	σ	c	m	s	$E(D)$
LGR	D_1	0.795	0.40	0.506	-0.344	0.478	0.709
LGS	D_2	0.775	0.41	0.524	-0.376	0.492	0.687
LMN	D_3	0.247	0.19	0.783	-1.638	0.692	0.194
MCN 1/	D_5	0.050	0.10	2.009	-3.808	1.271	0.022

1/ MCN D assumed to have non-zero mean and variance, chosen arbitrarily. Overall D values not sensitive to MCN D .

Using the FLUSH passage model in prospective mode, values for $P_{0,j}$ can be output for each water year under a particular scenario. An 'overall D ' value can be computed by weighting the D from each project by the portion of the transported population represented by that $P_{0,j}$:

$$\bar{D} = \sum_{j=1}^5 \frac{P_{0,j} \cdot D_j}{(1 - P_{0,n})}$$

Table F-4 shows the results of this exercise, using sequences of random D values drawn from distributions described above, for the 16 prospective water years (1977-92). Shown are the sensitivity to FGE assumption. Note from earlier equation that $P_{0,j}$'s don't depend on survival rates, so that Turb assumptions and predator reduction program assumptions are irrelevant (though they do matter to system survival). Also, because $P_{0,j}$'s are not dependent on passage model survival rates, results using CRiSP $P_{0,j}$'s would likely be similar to those from FLUSH.

Table F-4. Geometric Mean LGR D vs. Geometric 'Overall D ' from 16 year time series of random draws of D , for A1 and A2 with alternative assumptions about effectiveness of extended length screens, using Spring FLUSH $P_{0,j}$'s.

Scenario	FGE	Estimated LGR geomean D	Realized overall geomean D	Amount LGR D over-estimates
A1	Low	0.705	0.655	7.6%
A1	High	0.705	0.685	2.9%
A2	Low	0.705	0.636	11%
A2	High	0.705	0.689	2.3%

The results in Table F-4 suggest that a good estimate of LGR D (i.e. the geometric mean of a time series with either sufficiently low variance or a high number of observations) would be a reasonable approximation to overall D under the scenarios analyzed.

Several cautions are in order, however. This result is dependent on the findings so far that LGR D and LGS D are very similar. LGS D is important since the proportion destined to be transported at LGS ranges from about 16% to 29%, depending on scenario and FGE assumption. The assumptions about D at LMN and MCN have little effect, since the maximum average contribution of LMN is about 10%, and from MCN 5% or less (0% under A1).

Other factors that are not considered in the analysis that would affect the mean and variance of $P_{0,j}$'s are 1) inter- or intra-annual variance in Fish Guidance Efficiencies (FGEs); 2) inter- or intra-annual variance in spill effectiveness; 3) inter- or intra-annual variance in collection system survival rates. It also does not include co-variance between project-specific D 's from year to year, which, if extant, would result in greater overall variance in system survival and overall D . Further, it assumes that the relatively constant spill proportions within a scenario that come from the hydroregulation models used to do passage modeling accurately reflect the flow and spill management that would actually occur in the future.

Under the assumption that LGR D is a reasonable index for overall D , we can look at the combined effect of sampling and process variance on the ability to estimate the quantity of interest, geometric mean D from a series of annual estimates, with desired confidence. If variance in $\ln(D)$ is a result of both measurement (sampling) error and variation in environmental processes affecting SARs, then future observations of $\ln(D)$ can be considered as independent random draws from a normal distribution with this variance. Under these assumptions, the expected standard error of the future mean $\ln(D)$ estimates is equal to the square root of the variance in $\ln(D)$ divided by the square root of the number of years sampled. If n years of observed D 's are available, then the level of confidence that the observed mean $\ln(D)$ is greater than a desired mean $\ln(D)$ can be calculated by determining the cumulative probability from the normal distribution that an observed mean $\ln(D)$ over n years comes from a distribution where the true mean $\ln(D)$ is equal to the desired value and variance is equal to that observed to date. The normal distribution tested against would have mean $\ln(D) = \mu$ and standard error $= \sigma / \sqrt{n}$, where μ = the average $\ln(D)$ and σ = standard deviation of $\ln(D)$ observed to date. Conversely, the mean $\ln(D)$ that would have to be observed for a variable number of years to achieve a desired confidence that the true $\ln(D)$ is greater than the target level can be computed.

We can take the logarithm of the observed, weighted geometric mean D at LGR as a sample target and explore the influence of future observed means and number of data points on the amount confidence in rejecting a null hypothesis that overall D is less than the target value. The 0.705 geomean LGR D corresponds with an overall geomean D of about 0.65 (Table 4), so we can compare these results directly with those of Section 3.1, which assumes constant SARs and experimental sampling error only. Section A.1 is also focused only on detecting the relative likelihood of alternatives to null hypotheses on very different values of D (i.e. comparing $D \leq .35$ to $D \leq .65$), which probably aren't particularly relevant to determining the prospects for survival and recovery, since in PATH those prospects were less than desired even using models where D was drawn from a distribution with mean $\cong 0.65$.

Figure F-1 shows the confidence level versus number of years of data, for several different future observed geomean D values. From Figure F-1, it can be seen that if the future observed geometric mean D at LGR was 0.75, even after 75 years we would be less than 90% confident that true overall D was greater than the hypothesized value of 0.65. After 40 years, we would expect about 80% confidence that the true value exceeded the hypothesized value. The observed D would have to be 0.80 to expect 95% confidence after 40 years. In other words, we would have to observe a geometric mean LGR D of 0.8 over 40 years to have a 5% or less probability of incorrectly rejecting the null hypothesis that true overall D is less than or equal to 0.65.

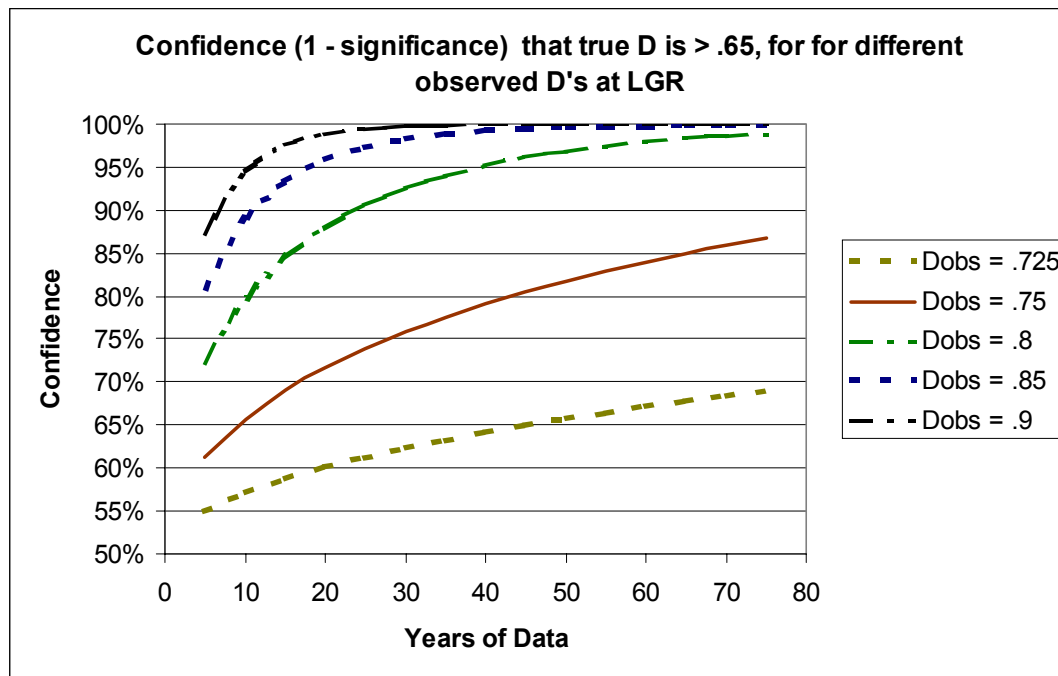


Figure F-1. Confidence level that true D is > 0.705 at LGR (overall D is $> \sim .65$) for different future observed geometric means of D for a time series of given length.

Figure F-2 shows the observed geometric mean D values needed for different levels of confidence after a given number of years. The analysis shown in Figures F-1 and F-2 can be performed for different target D levels, and for different observed D geomean and variance values resulting from different interpretations and weighting of the PIT-tag data (for some of the possibilities, see Bouwes et al. 1999). However, the present analysis suggests that under status quo scenarios, it will likely take many more years to determine with high or even moderate confidence whether the true future D value will be sufficient to give the Snake River stocks an acceptably high probability of survival and recovery than it would take to simply determine whether D was closer to 0.35 or 0.65. A high D value alone would not necessarily indicate that there is a high chance of survival and recovery under transportation-based options; see Bouwes et al. (1999) for other necessary assumptions. The risk of extinction to the stocks is, of course, positively related to the number of years of delay before taking meaningful action to improve population growth rates.

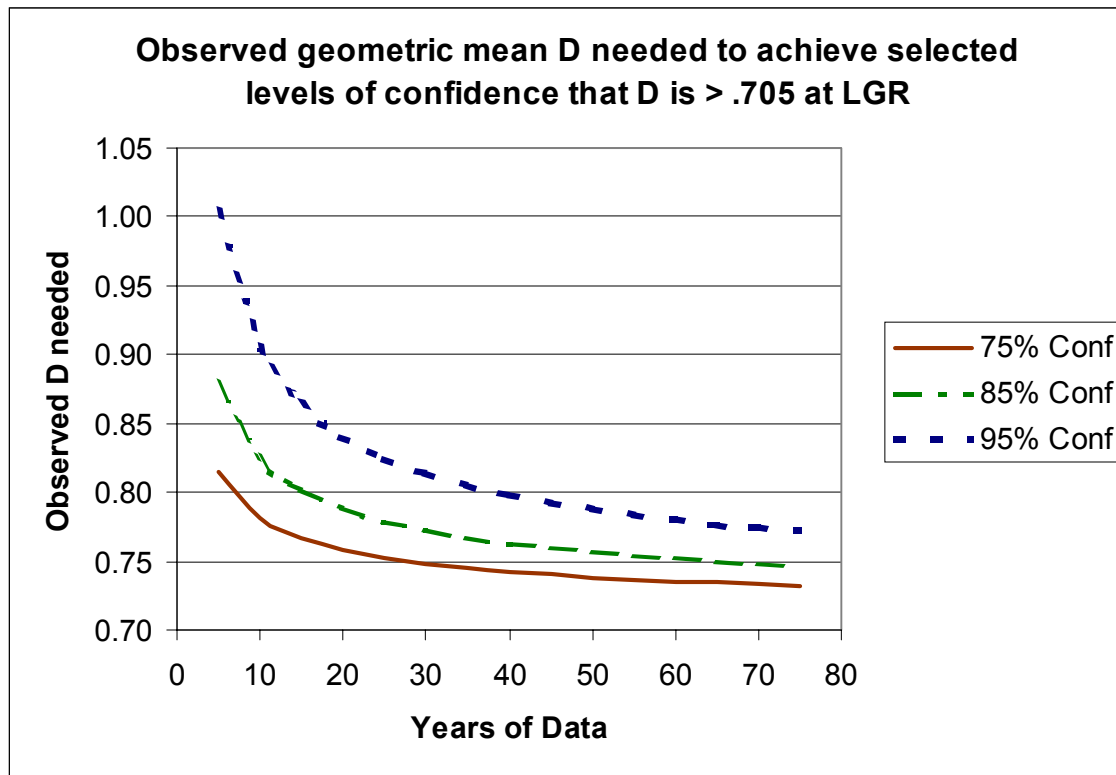


Figure F-2. Needed geometric mean observed D to achieve different levels of confidence that LGR D is > 0.705 (overall D is > ~ .65), for data sets of different length.

References

- Bouwes, N., H. Schaller, P. Budy, C. Petrosky, R. Kiefer, P. Wilson, O. Langness, E. Weber, and E. Tinus. 1999. An Analysis of Differential Delayed Mortality Experienced by Stream-type Chinook Salmon of the Snake River: Response to A-Fish Appendix. October 4, 1999
- Burgman, M.A., S. Ferson, and H.R. Açkakaya. 1993. Risk Assessment in Conservation Biology. Chapman & Hall, London. 314 p.
- Harmon, J.R. and 5 others. 1993. Research related to transportation of juvenile salmonids on the Columbia and Snake Rivers, 1992, Northwest Fisheries Science Center, NMFS, Seattle. October 1993.
- ISAB. 1999. Review of the National Marine Fisheries Service draft cumulative risk analysis addendum. ISAB Report 99-7, November 8, 1999.
- Zar, J.H. 1984. Biostatistical Analysis. Second edition. Prentice-Hall, Englewood Cliffs NJ. 718 p.

Appendix G. Application of PATH retrospective analysis to assumptions in the stream fertilization experiment

A limitation of the stream fertilization management experiment is that the PATH retrospective analysis does not provide any evidence of a temporal decrease in survival rate through the freshwater life stage that is proposed as the response variable in the experiments. Parr-to-smolt survival rates would be estimated for populations from streams with and without fertilization. However, the PATH retrospective analysis indicates that while life-cycle survival rates and SARs decreased after completion of the hydrosystem, there was little evidence of decreased survival rates through the freshwater spawning/rearing life stage. The decrease in spawner-to-smolt survival rates (residuals from ANCOVA Ricker fit) after 1975, if any, was not of a magnitude to explain the drastic decline in survival rates from spawner-adult recruits (Petrosky and Schaller 1996; PATH FY96 Retrospective Conclusions). Smolt-to-adult survival rates (SAR) decreased dramatically after 1970 (Petrosky and Schaller 1998), similar to declines evident in overall life-cycle survival rates. SARs and smolt/spawner estimates for smolt years 1964-1994 are shown in Figure G-1.

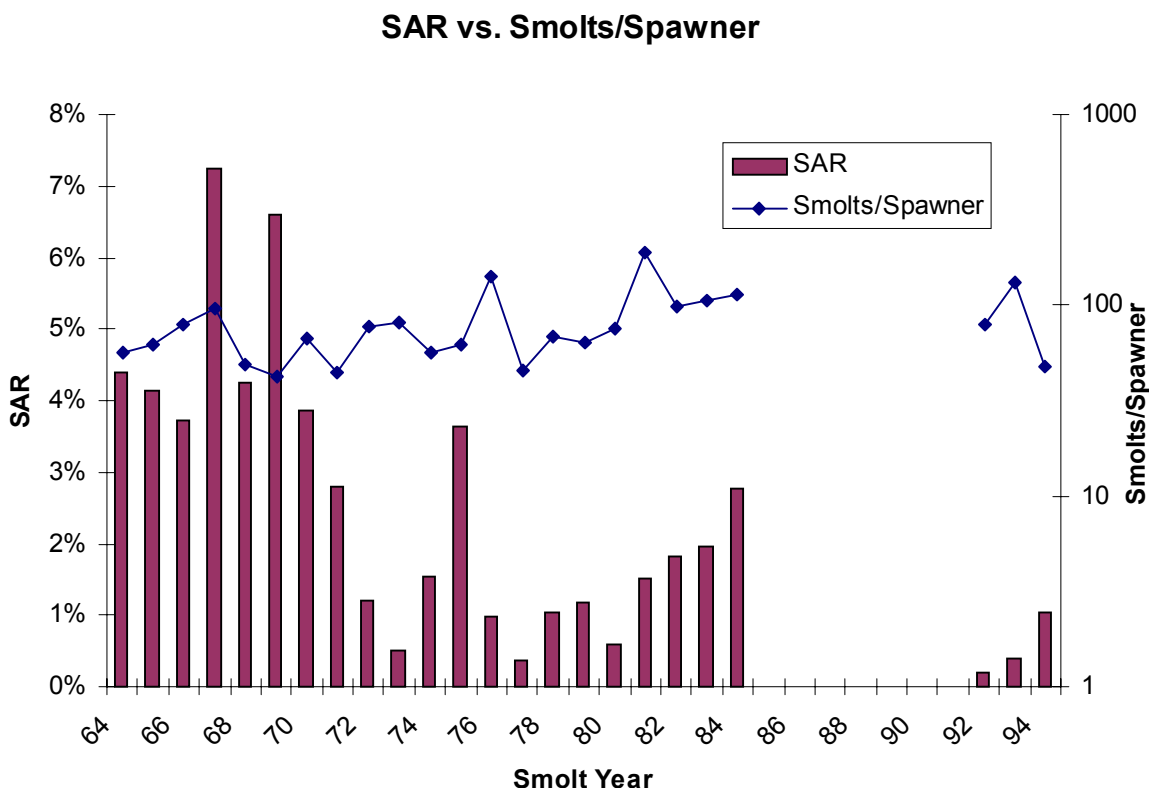


Figure G-1. Patterns of SAR and smolts/spawner (natural log scale) for Snake River wild spring/summer chinook, smolt years 1962-1994. Smolt/spawner estimates represented by SP1 and FGE=0.56 assumptions.

The fertilization experiments propose a potential reduction in total life cycle mortality ('m') for treatment populations through improvements in parr-smolt survival rates (described in section A.6.5). However, spawner-smolt survival rate (which includes that life stage) does not appear to be a good predictor of SAR (Figure G-2). Therefore, because the number of smolts produced per spawner did not decrease when the number of adult returns dramatically decreased (which occurred when smolt-to-adult survival rates

dramatically decreased) it seems unlikely that increases in carcass introductions will substantially improve spawner-to-smolt survivals.

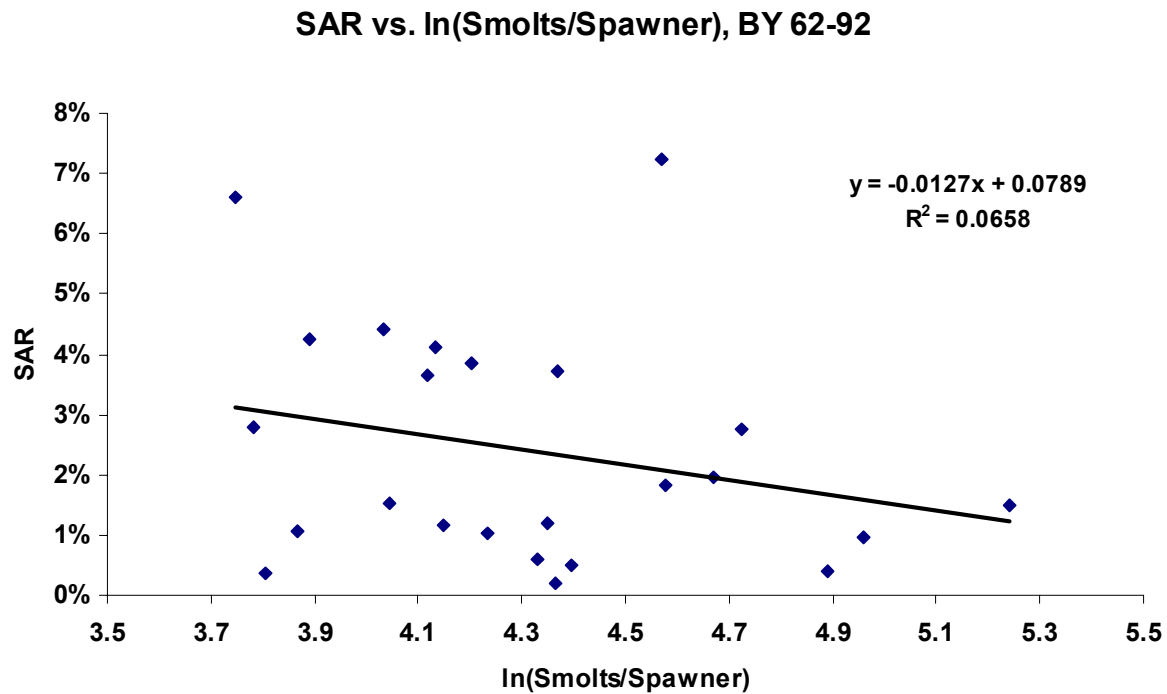


Figure G-2. Smolt-to-adult return rates versus ln(smolt/spawner), smolt years 1962-1992. The smolt/spawner index assumed SP1 and 0.56 FGE.

Appendix H. Experimental Management of D

These notes discuss how information on time varying SAR may be used to understand the processes contributing to the temporal varying nature of SAR and to assess if the transportation system can be optimized in light of the variations.

Mathematics of SAR and D

We begin by characterizing the smolt to adult return ratio (SAR) in terms of different distinct processes that make SAR different for transported and nontransported fish and which contribute to the SAR varying within the season and between seasons. The SAR for either transport or nontransport groups can be expressed in terms of four life stage survival factors: 1) F , a fish condition factor that we define as dependent on the arrival day or interval, x , of the fish to the transport dam, 2) E , an estuary/ocean condition factor that is a function of the estuary arrival day $y = x + d$ where d is the delay between arrival at the transport/release dam and arrival in Bonneville Dam tailrace, 3) H , a post Bonneville condition factor that depends on the route of passage, and 4) V , the passage survival in transport or inriver passage and may depend on time.

Although these factors are components of the total survival from smolt to adult, there is not necessarily a direct relationship between where the effect takes place and where the mortality occurs. In particular, mortality associated with F and H occur after the hydrosystem while events creating the mortality depend on time and passage route through the hydrosystem. For V , the hydrosystem passage, and the hydrosystem mortality are coincident. For E the timing of estuary passage and the estuary mortality are essentially coincident. Thus, the factor F characterizes how fish condition at the top of the hydrosystem affects survival below the hydrosystem, presumably in the estuary. The estuary factor E characterizes the effects of the estuary and the coastal ocean on fish survival moving through this habitat. E and F are common to both transport and nontransport fish and are distinguished by the specific times fish are at the top and the bottom of the hydrosystem as indicated by time intervals x and y . The hydrosystem factors V and H characterize the direct and delayed effects of the hydrosystem passage. The SAR, can be expressed as:

$$SAR_i(x, y) = F(x) E(y) H_i V_i(x) \quad (1)$$

Each passage route has a one-to-one correspondence between x and y such that

$$y_n = x + d_n \quad \text{and} \quad y_t = x + d_t \quad (2)$$

where d_n and d_t are the time for fish to travel from the transport dam to Bonneville Dam tailrace in river and in transportation.

The yearly averaged SAR for all fish passing a transport dam, including those transported and those passing inriver, can be expressed in terms of x and d as

$$\overline{SAR} = \sum (SAR_t(x, x + d_t) f(x) + SAR_n(x, x + d_n) (1 - f(x)) h(x) \quad (3)$$

where $f(x)$ is the fraction of the daily passage that is transported on day x , where $h(x)$ is the fraction of the seasonal dam passage on day x .

The ratio of the transport to control SARs referenced to the release date, x , characterizes the ratios of the time dependent effects of the estuary, the delayed mortality associated with passage, and the direct mortality of passage. In the transport to control ratio fish condition terms, $F(x)$ cancel so the TC ratio is

$$TC(x) = \frac{SAR_t(x, y_t)}{SAR_n(x, y_n)} = \frac{E(y_t)H_tV_t(x)}{E(y_n)H_nV_n(x)} \quad (4).$$

A Transport to Control ratio *adjusted* so the transport and control fish enter the estuary together, y , characterizes the difference in the condition of transport and nontransport groups and the effects of the estuary cancel since the two groups pass through the estuary together. The equation is:

$$TCA(y) = \frac{SAR_t(x, y)}{SAR_n(x - d_n + d_t, y)} = \frac{F(x)H_tV_t(x)}{F(x - d_n + d_t)H_nV_n(x)} \quad (5)$$

The differential delayed mortality between transported and nontransported fish defined on an interval basis using eq(4) is:

$$D(x) = \frac{SAR_t(x, y_t)}{SAR_n(x, y_n)} \frac{V_n(x)}{V_t(x)} = \frac{E(y_t)H_t}{E(y_n)H_n} \quad (6)$$

In a similar manner a D for fish from the adjusted transport and nontransport groups migrating through the estuary together can be expressed using eq(5).

$$DA(y) = \frac{SAR_t(x, y)V_n(x - d_n + d_t)}{SAR_n(x - d_n + d_t, y)V_t(x)} = \frac{F(x)H_t}{F(x - d_n + d_t)H_n} \quad (7)$$

$DA(y)$ is a measure of differential survival of transport and nontransported fish migrating through the estuary together. As such, both groups experience the same estuary survival factors and so this term cancels in the ratio. What is left measures the difference in the estuary survival resulting from the fish having arrived at Lower Granite Dam at different times and having arrived below Bonneville Dam by different passage routes. F and H characterize these factors. When $DA(y)$ is greater than 1 transport fish survival survive better and when it is less than 1 the inriver fish survive better.

Over a portion of the run the later part of the transport group, x_1 , and the early part of the inriver group, x_2 , travel through the estuary together so $y_1 \sim y_2$. Taking the ratio of D to DA for intervals over which $y_1 \sim y_2$ approximately holds we eliminate the passage specific delayed mortality factor as defined by H . What remains is the ratio of the estuary and fish conditions factors as defined

$$R(x_2, y_1) = \frac{D(x_2)}{DA(y_1)} = \frac{E(y_{2,t})}{E(y_{2,n})} \frac{F(x_1 - d_n + d_t)}{F(x_1)} \quad (8)$$

The representative seasonal D for transport from a dam is

$$D = \sum D(x)h(x) \quad (9)$$

where $h(x)$ is the fraction of the total run that passes the transport dam within an interval. A characteristic value of R can be expressed by weighting the interval values by the average fraction of the total run represented by each x_2, y_1 pair corresponding to release indexes x_2, x_1 . This becomes:

$$R = \frac{\sum R(x_2, y_1)(h(x_2) + h(x_1))}{\sum (h(x_2) + h(x_1))} \quad (10)$$

Data

To evaluate the impacts of SAR we use the 1995 PIT tag data. In this year PIT tag spring chinook passed Lower Granite Dam in large numbers as juveniles and significant numbers were recovered as adults at Lower Granite Dam. Fish passing LGR were put in barges and transported to below BON Dam or they were bypassed back into the river. From these data we are able to construct SAR and the other measures defined in the equations above. In practice because of the low tagging in some days and the low overall SAR some daily tagging groups had no adult returns. To alleviate this problem we group five days of tagging into a release group and thus characterized the SAR over five day intervals. The resulting SARs for the transport and control fish are illustrated in Figure H-1 below. In general the SAR for the transported fish increases with time while the SAR for the in-river passing fish decrease with time.

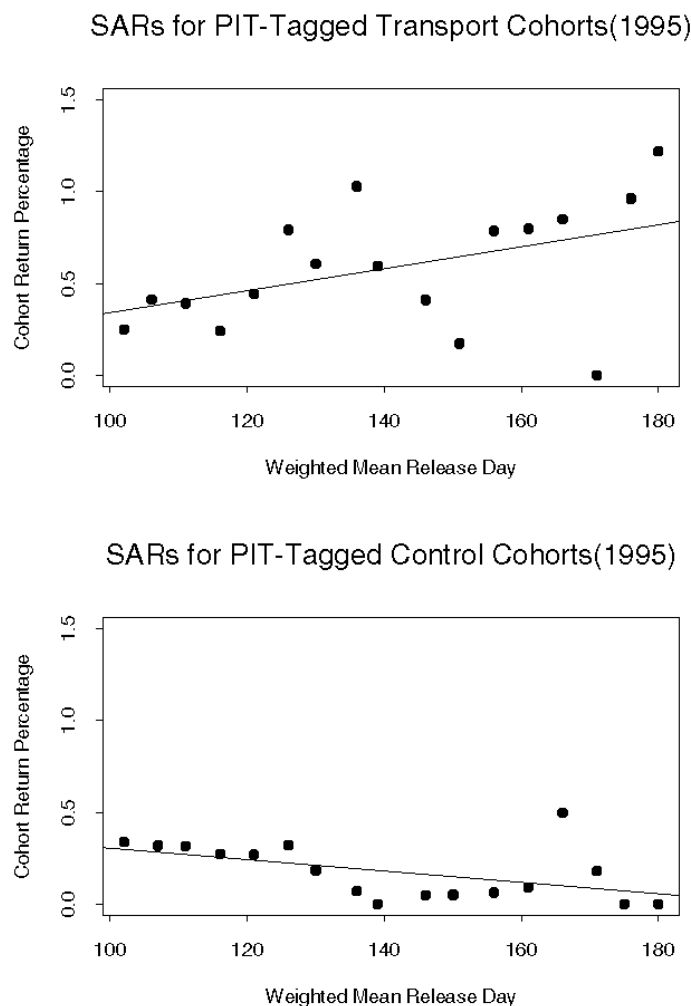


Figure H-1. SAR for transport and inriver passing fish for 5 day intervals.

To characterize the fractional daily passage distribution, h , we constructed run distribution based on the spring chinook passage index at LGR dam for 1995. This information was obtained from DART (Figure H-2). To characterize fractional percent of the total run transported each day, f , we used the CRiSP 1.6 passage model to characterize the percent of fish transported each day. Fish arrival time to below Bonneville dam was also characterized by the CRiSP 1.6 passage model.

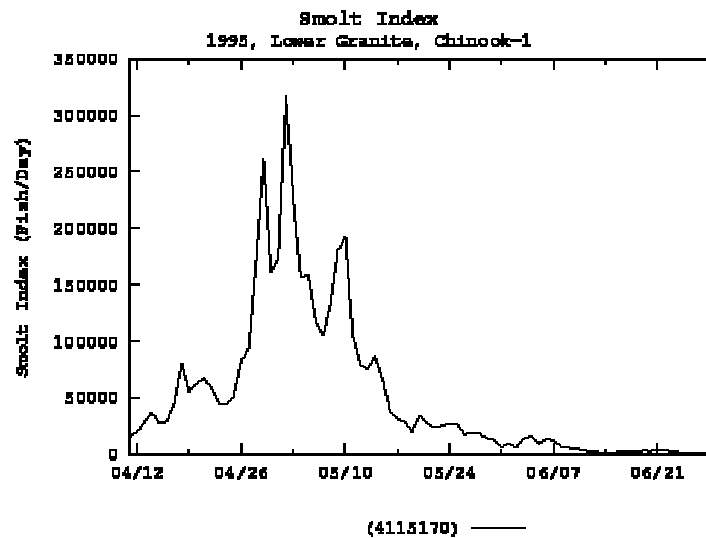


Figure H-2 Yearling chinook Smolt passage index at LGR dam in 1995.

The temporal pattern in $D(x)$ determined from eq(6) is illustrated in Figure H-3.

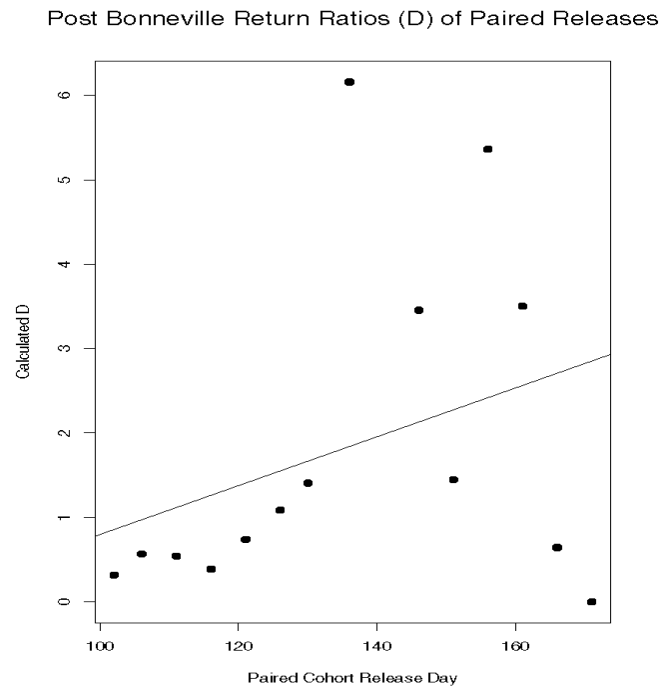


Figure H-3. Yearling chinook D varies over season.

The observations used in the analysis are given in the Table H-1 below. SAR intervals represent 5 days of daily SAR from the 1995 PIT tag studies.

Table H-1. SAR data used to estimate parameters.

Mean release day (x)	Control Release		Transport Release	
	Release Total	SAR	Release Total	SAR
102	2668	0.337	3197	0.250
106	10378	0.318	3161	0.411
111	22167	0.316	20420	0.392
116	19112	0.272	16262	0.240
121	28984	0.269	25729	0.443
126	29885	0.321	19849	0.791
130	10960	0.183	6930	0.606
136	4362	0.069	4180	1.029
139	848	0.000	842	0.594
146	2054	0.049	1710	0.409
151	2003	0.050	1735	0.173
156	1625	0.061	1145	0.786
161	1093	0.091	876	0.799
166	403	0.496	235	0.851
171	556	0.180	520	0.000
176	407	0.000	208	0.962
180	169	0.000	164	1.219

The calculated values y_n , y_t , V_n , V_t , $D(x)$, $f(x)$ and $h(x)$ are given in Table H-2. The estimated arrival times at Bonneville Dam are developed from the CRiSP 1.6 passage model. Also the average value of D between the first interval day 102 and interval x is given in $E\{D(x)\}$. In Table H-3 the adjusted measure $DA(y)$ for each interval and average values up through an interval $E\{D(y)\}$ plus the intervals used to calculate $D(y)$ are given. Also shown are the $D(x_2)$ values and the calculation of R .

The representative value of R as determined by eq(10) is 1.19. The average value of D determined from eq(9) is 1.35 and the representative value of DA is 1.5. In all these measures the transported group have higher survival than the inriver group. Furthermore, the D value representing the effect of estuary survival increases over the season (Figure H-3).

Table H-2. Calculated values include the inriver survival V_n , transport survival V_t , fraction of the run arriving in the interval h , fraction of the interval that was transported f , interval value of $D(x)$, and $E\{D(x)\}$ which is the average value of D from the interval 102 to interval x .

Mean release day (x)	y_n	y_t	V_n	V_t	f	h	$D(x)$	$E\{D(x)\}$
102	136	104	0.414	0.97	0.018	0.552	0.32	0.320
106	137	106	0.425	0.97	0.052	0.551	0.57	0.506
111	139	113	0.425	0.97	0.069	0.546	0.54	0.523
116	141	118	0.429	0.97	0.107	0.576	0.39	0.465
121	144	123	0.436	0.97	0.276	0.450	0.74	0.610
126	148	128	0.428	0.97	0.162	0.395	1.09	0.724
130	152	132	0.411	0.97	0.153	0.550	1.41	0.849
136	158	138	0.400	0.97	0.061	0.437	6.16	1.210
139	160	141	0.399	0.97	0.031	0.437	Inf	1.210
146	167	148	0.399	0.97	0.026	0.534	3.45	1.273
151	170	153	0.405	0.97	0.011	0.518	1.45	1.275
156	177	158	0.407	0.97	0.015	0.511	5.36	1.340
161	183	163	0.389	0.97	0.006	0.531	3.50	1.353
166	190	168	0.365	0.97	0.002	0.491	0.64	1.352
171	197	173	0.339	0.97	0.004	0.554	0.00	1.352
176	203	179	0.320	0.97	0.001	0.569	Inf	1.352
180	210	182	0.314	0.97	0.001	0.566	Inf	1.352

Table H-3. The adjusted differential delayed mortality DA according to eq(7) and R according to eq(8). Also shown are the data grouping intervals used to calculate DA .

BON Arrival Transport y_1	BON Arrival Control y_2	f for x_1	f for x_2	$DA(y_1)$	$D(x_2)$	$R(x_2, y_1)$
138	136	6.1	1.9	1.30	0.32	0.25
138	137	6.1	5.3	1.42	0.57	0.40
138	139	6.1	7.0	1.43	0.54	0.38
141	141	3.1	10.7	0.97	0.39	0.40
141	144	3.1	27.0	0.99	0.74	0.75
148	148	2.6	27.6	0.56	1.09	1.95
153	152	1.1	16.2	0.40	1.41	3.53
158	158	1.5	15.3	4.70	6.16	1.31
163	160	0.6	6.1	Inf		
168	167	0.2	3.1	7.19	3.45	0.48
168	170	0.2	2.6	7.11	1.45	0.20
173	170	0.4	1.1	0.00	1.45	
179	177	0.1	1.5	6.66	5.36	0.80
182	183	0.1	0.6	5.37	3.5	0.65

Evaluating actions to optimize transportation

To improve the effectiveness of transportation we can alter the time required for transported fish to reach the estuary, as defined by the factor d_t , or we can alter the daily fraction of transported fish as characterized by h . The relationship between these variables and the yearly averaged SAR is given in eq(3). To explore the impact of these two actions assume that the 1995 SARs for transported and inriver-passing fish characterize the estuary survival factor $E(y)$. Optimizing SAR then involves either the altering the arrival time of transport fish into the estuary, which changes d_t or by increasing the percent of fish that are transported, which changes h . A third option of delaying the beginning of the transport season uniformly lowers SAR and so it is not considered further. The effect of increasing the transport time, δ , is illustrated in Figure 4. The distribution of arrivals times at the transport dam $f(x)$ exhibits a peaked distribution. The Bonneville dam arrival time of transported fish is given by y_t and the distribution of fish with an additional delay is $y_t + \delta$. The displacement between x and y_t is the passage time, d , which currently is on the order of 2 days for transported fish. We consider delays of 5, 10 15 and 20 days. By delaying the transport time the fish are more likely to enter the estuary when the potential SAR is higher under the assumption that the SAR is determined by the estuary survival factor increasing with season.

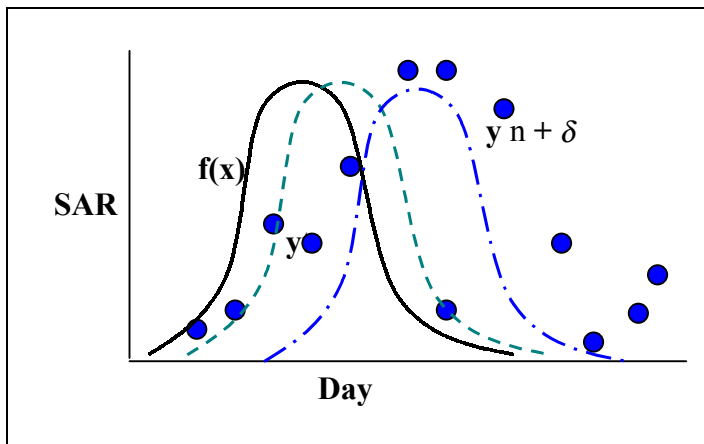


Figure H-4. Illustration of transport SAR (dots), transport dam arrival distribution $f(x)$, Bonneville arrival distribution y_t , and distribution with a delay of δ .

To explore effects of these two actions we can use the SAR distribution and the arrival time distribution of fish x for 1995. The SAR is referenced to the time at arrival to Bonneville Dam. We assume the changes in SAR are a result of estuary arrival timing. We then adjust d_t and F to alter the pattern over which fish enter the estuary. By these adjustments arriving fish experience SAR depend on when they arrive in the estuary and by which passage route they take. In this approach we have assumed that the distribution of SAR by the two passage routes are fixed. We take the 1995 transport experiment to represent the patterns. Our question then comes to “how would the average SAR for 1995 have been altered if we had moved fish at a different rate in transportation and if we had used a different transport schedule.

The impacts of slowing the barge transport by 5-day intervals on the overall SAR is illustrated in Figure H-5 below. The impact of altering the percent of fish transported on the total SAR is illustrated in Figure H-6.

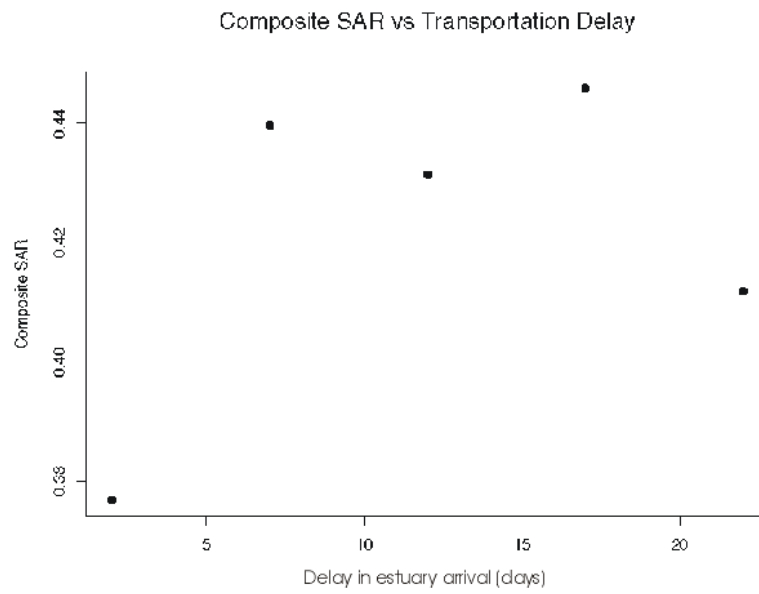


Figure H-5. SAR for delays in transport fish arrival Below Bonneville Dam.

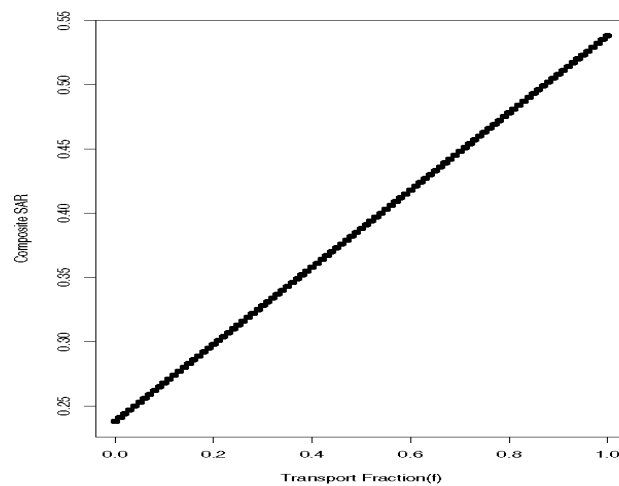


Figure H-6. Change in total SAR by altering the fraction of fish transported.

Appendix I. Details of Bayesian and Bootstrap Sampling

For the bayesian sampling, we did the following

- (1) Computed the maximum likelihood estimates of the parameters, $\hat{\beta}$, and the unscaled covariance matrix, V_{β} .
- (2) Computed the standard error the regression, s^2 .
- (3) Drew a σ^2 from the scaled inverse- χ^2 distribution, $\text{Inv-}\chi^2(n-k, s^2)$, where n is the number of observations and k is the number of parameters.
- (4) Drew a sample of the parameters from the multivariate normal distribution $N(\hat{\beta}, V_{\beta}\sigma^2)$.

Further details of this method may be found in Gelman et al. (1995), section 8.3 (pages 235-239).

For the bootstrap sampling, we:

- (1) Calculated the least squares predicted $\log(R/S)$ and the residuals of the regression.
- (2) Randomly selected n of the residuals (with replacement) and added them to the predicted $\ln(R_{i,t}/S_{i,t})$ to obtain a bootstrap responses, $\ln(R_{i,t}/S_{i,t})^*$.
- (3) The bootstrap responses $\ln(R_{i,t}/S_{i,t})^*$ are then substituted for the actual responses in a regression and the resulting parameter estimates, $\hat{\beta}^*$, represent the random sample from the parameter space.

Further details are found in Efron and Tibshirani 1993, section 9.4, pages 111-112.

Efron, B and R.J Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Appendix J. Implementation of the Hydro Extra mortality hypothesis

One of the hypotheses entertained by PATH was the hydro extra mortality hypothesis which stated that the change in productivity of the Snake River stocks not due to transportation, direct effects of passage mortality, or climate effects (in common with the downriver stocks) was due to some loss of productivity related specifically to the Snake River dams. The hypothesis states that if dams are removed, then this productivity will be restored. To estimate the loss of productivity due to extra mortality, we used the following equation:

$$\bar{m}_{1957-1974} - \bar{m}_{1978-1994} = (\bar{\delta}_{1957-1974} - \bar{\delta}_{1978-1994}) + (\bar{\omega}_{1957-1974} - \bar{\omega}_{1978-1994}) + (\ln(\bar{\lambda}_{n1957-1974}) - \ln(\bar{\lambda}_{n1978-1994}))$$

average change in m series between 1957-1974 and 1978-1994 periods	=	average change in delta series (common year effects) between 1957-1974 and 1978-1994 periods	+	average change in system survival (depends on D hypothesis) between 1957-1974 and 1978-1994 periods	+	average change in ln(post- Bonneville survival factor of in-river fish) between 1957- 1974 and 1978-1994 periods
---	---	---	---	---	---	---

The “extra mortality” is then given by the negative of the last term in parentheses: minus the change in ln(post-Bonneville survival factor of in-river fish). Since system survivals were not available from 1991-1994 and deltas were unavailable for 1991-1994, we used the 1978-1990 averages of these to approximate their 1978-1994 averages. By solving the equation for the change in extra mortality, we get the following equation

$$-extra_mortality = -(\bar{\delta}_{1957-1974} - \bar{\delta}_{1978-1994}) - (\bar{\omega}_{1957-1974} - \bar{\omega}_{1978-1994}) + (\bar{m}_{1957-1974} - \bar{m}_{1978-1994})$$

For the hydro hypothesis, this “extra mortality,” which depends on the D hypothesis (because the D hypothesis changes system survival), disappears when the Snake River dams are removed. Thus, prospectively, the Δm s will be adjusted upwards by $-extra_mortality$ when the Snake dams are removed (4-dam drawdown). This extra_mortality adjustment for the various D hypothesis is given in the table below.

D hypothesis	sys_surv1- sys_surv2	m1-m2	delta1-delta2	log(l1)-log(l2)
D=.3	0.4525438	1.278968	0.385258715	0.441165252
D=.6	-0.038514408	1.278968	0.385258715	0.93222346
D=.8	-0.252926033	1.278968	0.385258715	1.146635085

Appendix K. Comparison of normal approximation to the actual distribution of estimated Δm 's

We compared how well the $1.64 \times \text{stderr}$ approximates the true critical value of the Δm estimate. The $1.64 \times \text{stderr}$ approximation is based on estimating the true Δm sampling distribution with a normal distribution (Figure K-1). The figure shows the frequency distribution for Δm using 28 control years and 5 treatment years. Notice that the true frequency distribution is slightly skewed to the left. This means that the critical value estimate of $1.64 \times \text{stderr}$ will be slightly biased upward. Indeed, for this example, the actual critical value was 0.75 compared to the normal approximation ($1.64 \times \text{stderr}$) of 0.81.

The bottom line, is that the approximation is biased high due to the fact that the frequency distribution is skewed to the left (probably due to the low 1990 and 1991 year effects), and therefore the critical value sets the bar a little too high (i.e., by using the $\text{stderr} \times 1.64$ critical value, we are really setting the alpha at less than 0.05). For the above example, using $\text{stderr} \times 1.64$ sets the significance level at 0.039, slightly lower than 0.05. As the sample size for the treatment years increases, we expect the approximation to improve.

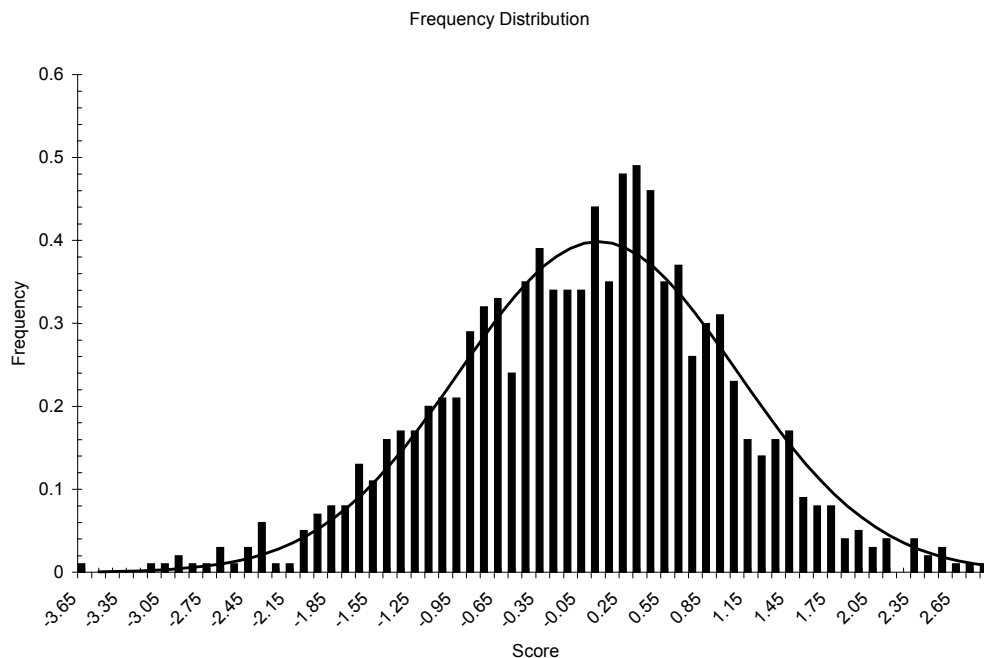


Figure K-1. Distribution of estimated Δm values (standardized to mean=0 and std. Dev.=1) for the 1/0 on/off generic experiment. The vertical bars represents the actual distribution; the smoothed curve is the normal approximation.